



(19)

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 939 128 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
01.09.1999 Bulletin 1999/35

(51) Int. Cl.⁶: **C12N 15/51**, **C07K 14/18**,
C12Q 1/68, **G01N 33/576**

(21) Application number: 99101746.8

(22) Date of filing: 17.09.1990

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IT LI LU NL SE

(30) Priority: 15.09.1989 US 408045
21.12.1989 US 456142

(62) Document number(s) of the earlier application(s) in
accordance with Art. 76 EPC:
90310149.1 / 0 419 182

(71) Applicants:
• Oya, Akira, Dr.,
Director General of the National Institute of
Health of Japan,
on behalf of The National Institute of Japan
Shinagawa-ku, Tokyo 141 (JP)
• CHIRON CORPORATION
Emeryville, California 94608 (US)

(72) Inventors:
• Miyamura, Tatsuo
Tokyo 168 (JP)
• Saito, Isumu
Tokyo 151 (JP)

- Houghton, Michael
Danville, California 94526 (US)
- Welner, Amy J.
Benicia, California 94510 (US)
- Han, Jang
Lafayette, California 94549 (US)
- Kolberg, Janice A.
Richmond, California 94804 (US)
- Cha, Tai-An
San Ramon, California 94583 (US)
- Irvine, Bruce Duncan
Concord, California 94519 (US)

(74) Representative:
Goldin, Douglas Michael
J.A. KEMP & CO.
14 South Square
Gray's Inn
London WC1R 5LX (GB)

Remarks:

This application was filed on 11 - 02 - 1999 as a
divisional application to the application mentioned
under INID code 62.

(54) **New HCV Isolates**

(57) Two new isolates of the Hepatitis C virus (HCV), J1 and J7, are disclosed. These new isolates comprise nucleotide and amino acid sequences which are distinct from the prototype HCV isolate, HCV1. Thus, J1 and J7 provide new polynucleotides and polypeptides for use, inter alia, in diagnostics, recombinant protein production and vaccine development.

EP 0 939 128 A2

DescriptionTechnical Field

- 5 [0001] The present invention relates to new isolates of the viral class Hepatitis C, polypeptides, polynucleotides and antibodies derived therefrom, as well as the use of such polypeptides, polynucleotides and antibodies in assays (e.g., immunoassays, nucleic acid hybridization assays, etc.) and in the production of viral polypeptides.

Background

- 10 [0002] Non-A, Non-B hepatitis (NANBH) is a transmissible disease or family of diseases that are believed to be viral-induced, and that are distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). NANBH was first identified in transfused
15 individuals. Transmission from man to chimpanzee and serial passage in chimpanzees provided evidence that NANBH is due to a transmissible infectious agent or agents. Epidemiologic evidence is suggestive that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, until recently, no transmissible agent responsible for NANBH had not been identified.

- [0003] Clinical diagnosis and identification of NANBH has been accomplished primarily by exclusion of other viral
20 markers. Among the methods used to detect putative NANBH antigens and antibodies are agar-gel diffusion, counter-immunoelectrophoresis, immunofluorescence microscopy, immune electron microscopy, radioimmunoassay, and enzyme-linked immunosorbent assay. However, none of these assays has proved to be sufficiently sensitive, specific, and reproducible to be used as a diagnostic test for NANBH.

- [0004] Until recently there has been neither clarity nor agreement as to the identity or specificity of the antigen antibody systems associated with agents of NANBH. It is possible that NANBH is caused by more than one infectious agent
25 and unclear what the serological assays detect in the serum of patients with NANBH.

- [0005] In the past, a number of candidate NANBH agents were postulated. See, e.g., Prince (1983) Ann. Rev. Microbiol. 37:217; Feinstone & Hoofnagle (1984) New Eng. J. Med. 311:185; Overby (1985) Curr. Hepatol. 5:49; Overby (1986) Curr. Hepatol. 6:65; Overby (1987) Curr. Hepatol. 7:35; and Iwarson (1987) British Med. J. 295:946. However, there
30 is no proof that any of these candidates represent the etiological agent of NANBH.

- [0006] In 1987, Houghton et al. cloned the first virus definitively linked to NANBH. See, e.g., EPO Pub. No. 318,216; Houghton et al., Science 244:359 (1989). Houghton et al. described therein the cloning of an isolate from a new viral class, hepatitis C virus (HCV), the prototype isolate described therein being named "HCV1". HCV is a Flavi-like virus,
35 with an RNA genome. Houghton et al. described the production of recombinant proteins from HCV sequences that are useful as diagnostic reagents, as well as polynucleotides useful in diagnostic hybridization assays and in the cloning of additional HCV isolates.

- [0007] The demand for sensitive, specific methods for screening and identifying carriers of NANBH and NANBH contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and NANBH accounts for up to 90% of these cases. There is a frequent progression to chronic liver
40 damage (25-55%).

- [0008] Patient care as well as the prevention of transmission of NANBH by blood and blood products or by close personal contact require reliable diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to NANBH. In addition, there is also a need for effective vaccines and immunotherapeutic therapeutic agents for the prevention and/or treatment of the disease.

- 45 [0009] While at least one HCV isolate has been identified which is useful in meeting the above needs, additional isolates, particularly those with divergent a genome, may prove to have unique applications.

Summary of the Invention

- 50 [0010] New isolates of HCV has been characterized from Japanese blood donors who have been implicated as NANBH carriers. These isolates exhibit nucleotide and amino acid sequence heterogeneity with respect to the prototype isolate, HCV1, in several viral domains. It is believed that these distinct sequences are of importance, particularly in diagnostic assays and in vaccine development.

- [0011] In one embodiment, the present invention provides a DNA molecule comprising a nucleotide sequence of at
55 least 15 bp from an HCV isolate substantially homologous to an isolate selected from the group J1 or J7, wherein said nucleotide sequence is distinct from the nucleotide sequence of HCV isolate HCV1.

- [0012] In another embodiment, the present invention provides a DNA molecule comprising a nucleotide sequence of at least 15 bp encoding an amino acid sequence from a HCV isolate J1 or J7 wherein the J1 or J7 amino acid sequence

is distinct from the amino acid sequence of HCV isolate HCV1.

[0013] Yet another embodiment of the present invention provides a purified polypeptide comprising an amino acid sequence from an HCV isolate substantially homologous to an isolate selected from the group J1 and J7, wherein said amino acid sequence is distinct from the sequence of the polypeptides encoded by the HCV isolate HCV1.

5 [0014] Still another embodiment of the present invention provides a polypeptide comprising an amino acid sequence from a HCV isolate J1 or J7 wherein the J1 or J7 amino acid sequence is distinct from the amino acid sequence of HCV isolate HCV1 and the polypeptide is immobilized on a solid support.

[0015] In a further embodiment of the present invention, an immunoassay for detecting the presence of anti-HCV antibodies in a test sample is provided comprising: (a) incubating the test sample under conditions that allow the formation
10 of antigen-antibody complexes with an immunogenic polypeptide comprising an amino acid sequence from an HCV isolate substantially homologous to an isolate selected from the group J1 and J7, wherein the amino acid sequence is distinct from the amino acid sequence of HCV isolate HCV1; and (b) detecting an antigen-antibody complex comprising the immunogenic polypeptide.

[0016] The present invention also provides a composition comprising anti-HCV antibodies that bind an HCV epitope substantially free of antibodies that do not bind an HCV epitope, wherein: (a) the HCV epitope comprises an amino acid
15 sequence from an HCV isolate substantially homologous to an isolate selected from the group J1 and J7, wherein the amino acid sequence is distinct from the amino acid sequence of HCV isolate HCV1; and (b) the J1 or J7 amino acid sequence is not immunologically cross-reactive with HCV1.

[0017] A further embodiment of the present invention provides an immunoassay for detecting the presence of an HCV polypeptide in a test sample comprising: (a) incubating the test sample under conditions that allow the formation of anti-
20 gen-antibody complexes with anti-HCV antibodies that bind an HCV epitope wherein: (i) the HCV epitope comprises an amino acid sequence from a HCV isolate J1 or J7; (ii) the J1 or J7 amino acid sequence is distinct from the amino acid sequence of HCV isolate HCV1; and (iii) the J1 or J7 amino acid sequence is not immunologically cross-reactive with HCV1; and (b) detecting an antigen-antibody complex comprising the anti-HCV antibodies.

25 [0018] Also provided by the present invention is a method of producing anti-HCV antibodies comprising administering to a mammal a polypeptide comprising an amino acid sequence from a HCV isolate J1 or J7 wherein the J1 or J7 amino acid sequence is distinct from the amino acid sequence of HCV isolate HCV1 whereby the mammal produces anti-HCV antibodies.

[0019] Yet another embodiment of the present invention provides a method of detecting HCV polynucleotides in a test
30 sample comprising: (a) providing a probe comprising the DNA molecule of claim 1; (b) contacting the test sample and the probe under conditions that allow for the formation of a polynucleotide duplex between the probe and its complement in the absence of substantial polynucleotide duplex formation between the probe and non-HCV polynucleotide sequences present in the test sample; and (c) detecting any polynucleotide duplexes comprising the probe.

[0020] A still further embodiment of the present invention provides a method of producing a recombinant polypeptide
35 comprising an HCV amino acid sequence, the method comprising: (a) providing host cells transformed by a DNA construct comprising a control sequences for the host cell operably linked to a coding sequence encoding an amino acid sequence from a HCV isolate J1 or J7 wherein the J1 or J7 amino acid sequence is distinct from the amino acid sequence of HCV isolate HCV1; (b) growing the host cells under conditions whereby the coding sequence transcribed and translated into the recombinant polypeptide; and (c) recovering the recombinant polypeptide.

40 [0021] These and other embodiments of the present invention will be readily apparent to those of ordinary skill in the art in view of the following description.

Brief Description of the Figures

45 [0022]

Figure 1 shows the consensus sequence of the coding strand of a fragment from the J7 C/E domain with the heterogeneities.

50 Figure 2 shows the consensus sequence of the coding strand of a fragment from the J1 E domain with the heterogeneities.

Figure 3 shows the consensus sequence of the coding strand of a fragment of the J1 E/NS1 domain with the heterogeneities.

Figure 4 shows the consensus sequence of the coding strand of a fragment from the J1 NS3 domain with the heterogeneities.

55 Figure 5 shows the consensus sequence of the coding strand of a fragment from the J1 NS5 domain with the heterogeneities.

Figure 6 shows the homology of the J7 C/E consensus sequence with the nucleotide sequence of the same domain from HCV1.

Figure 7 shows the homology of the J1 E consensus sequence with the nucleotide sequence of the same domain from HCV1.

Figure 8 shows the homology of the J1 E/NS1 consensus sequence with the nucleotide sequence of the same domain from HCV1.

5 Figure 9 shows the homology of the J1 NS3 consensus sequence with the nucleotide sequence of the same domain from HCV1.

Figure 10 shows the homology of the J1 NS5 consensus sequence with the nucleotide sequence of the same domain from HCV1.

Figure 11 shows the putative genomic organization of the HCV1 genome.

10 Figure 12 shows the nucleotide sequence of the ORF of HCV1. In the figure nucleotide number 1 is the first A of the putative initiating methionine of the large ORF; nucleotides upstream of this nucleotide are numbered with negative numbers.

Figure 13 shows the consensus sequence of the coding strand of a fragment from the J1 NS1 domain (J1 1519) with the nucleotide sequence of the same domain from HCV1. Also shown are the amino acids encoded therein.

15 Figure 14 shows a composite of the consensus sequence from the core to the NS1 domain of J1 with the nucleotide sequence of the same domain from HCV1. Also shown are the amino acids encoded therein.

Figure 15 shows a consensus sequence of the coding strand of the NS1 domain of J1, as determined in Example IV. Also shown are the nucleotide sequence of the same domain from HCV1, and the amino acids encoded in the HCV1 and J1 sequences.

20 Figure 16 shows a consensus sequence of a coding strand of the C200 region of the NS3-NS4 domain of J1. Also shown are the nucleotide sequence of the same domain from HCV1. Also shown are the amino acids encoded in the sequences.

Figure 17 shows a consensus sequence of the coding strand of the NS1 domain of J1, as determined in Example V. Also shown are the nucleotide sequence of the same domain from HCV1, and the amino acids encoded in the sequences.

25 Figure 18 shows a consensus sequence of the coding strand of the untranslated and core domains of J1. Also shown are the nucleotide sequence of the same domain from HCV1, and the amino acids encoded in the sequences.

30 Detailed Description of the Invention

[0023] The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA techniques, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fritsch & Sambrook, MOLECULAR CLONING; 35 A LABORATORY MANUAL (1982); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS 40 FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent 45 applications, and other publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

[0024] The term "hepatitis C virus" has been reserved by workers in the field for an heretofore unknown etiologic agent of NANBH. Accordingly, as used herein, "hepatitis C virus" (HCV) refers to an agent causative of NANBH, which was formerly referred to as NANBV and/or BB-NANBV from the class of the prototype isolate, HCV1, described by 50 Houghton et al. See, e.g., EPO Pub. No. 318,216 and U.S. Patent App. Serial No. 355,002, filed 19 May 1989 (available in non-U.S. applications claiming priority therefrom), the disclosures of which are incorporated herein by reference. The nucleotide sequence and putative amino acid sequence of HCV1 is shown in Figure 6. The terms HCV, NANBV, and BB-NANBV are used interchangeably herein. As an extension of this terminology, the disease caused by HCV, formerly called NANB hepatitis (NANBH), is called hepatitis C. The terms NANBH and hepatitis C may be used interchangeably 55 herein. The term "HCV", as used herein, denotes a viral species of which pathogenic strains cause NANBH, as well as attenuated strains or defective interfering particles derived therefrom.

[0025] HCV is a Flavi-like virus. The morphology and composition of Flavivirus particles are known, and are discussed by Brinton (1986) THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and

Wagner, vol eds. Schlesinger and Schlesinger, Plenum Press), p.327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter.

5 [0026] The HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, i.e., reportedly on the order of 10^{-3} to 10^{-4} per incorporated nucleotide. Therefore, there are multiple strains, which may be virulent or avirulent, within the HCV class or species.

[0027] It is believed that the genome of HCV isolates is comprised of a single ORF of approximately 9,000 nucleotides to approximately 12,000 nucleotides, encoding a polypeptide similar in size to that of HCV1, an encoded polypeptide of 10 similar hydrophobic and antigenic character to that of HCV1, and the presence of co-linear peptide sequences that are conserved with HCV1. In addition, the genome is believed to be a positive-stranded RNA.

[0028] Isolates of HCV comprise epitopes that are immunologically cross-reactive with epitopes in the HCV1 genome. At least some of these are epitopes unique to HCV when compared to other known Flaviviruses. The uniqueness of the epitope may be determined by its immunological reactivity with anti-HCV antibodies and lack of immunological reactivity 15 with antibodies to other Flavivirus species. Methods for determining immunological reactivity are known in the art, for example, by radioimmunoassay, by ELISA assay, by hemagglutination, and several examples of suitable techniques for assays are provided herein.

[0029] It is also expected that the overall homology of HCV isolates and HCV1 genomes at the nucleotide level probably will be about 40% or greater, probably about 60% or greater, and even more probably about 80% to about 90% or 20 greater. In addition that there are many corresponding contiguous sequences of at least about 13 nucleotides that are fully homologous. The correspondence between the sequence from a new isolate and the HCV1 sequence can be determined by techniques known in the art. For example, they can be determined by a direct comparison of the sequence information of the polynucleotide from the new isolate and HCV1 sequences. Alternatively, homology can be determined by hybridization of the polynucleotides under conditions which form stable duplexes between homologous 25 regions (for example, those which would be used prior to S_1 digestion), followed by digestion with single-stranded specific nuclease(s), followed by size determination of the digested fragments.

[0030] Because of the evolutionary relationship of the strains or isolates of HCV, putative HCV strains or isolates are identifiable by their homology at the polypeptide level. Thus, new HCV isolates are expected to be more than about 40% homologous, probably more than about 70% homologous, and even more probably more than about 80% homologous, 30 and possibly even more than about 90% homologous at the polypeptide level. The techniques for determining amino acid sequence homology are known in the art. For example, the amino acid sequence may be determined directly and compared to the sequences provided herein. Alternatively the nucleotide sequence of the genomic material of the putative HCV may be determined, the amino acid sequence encoded therein can be determined, and the corresponding regions compared.

35 [0031] The ORF of HCV1 is shown in Figure 12. The non-structural, core, and envelope domains of the polypeptide have been predicted for HCV1 (Figure 5). The "C", or core, polypeptide is believed to be encoded from the 5' terminus to about nucleotide 345 of HCV1. The putative "E", or envelope, domain of HCV1 is believed to be encoded from about nucleotide 346 to about nucleotide 1050. Putative NS1, or non-structural one domain, is thought to be encoded from about nucleotide 1051 to about nucleotide 1953. For the remaining domains, putative NS2 is thought to be encoded 40 from about nucleotide 1954 to about nucleotide 3018, putative NS3 from about nucleotide 3019 to about nucleotide 4950, putative NS4 from about nucleotide 4951 to about nucleotide 6297, and putative NS5 from about nucleotide 6298 to the 3' terminus respectively. The above boundaries are approximations based on an analysis of the ORF. The exact boundaries can be determined by those skilled in the art in view of the disclosure herein.

[0032] "HCV/J1" or "J1" and "HCV/J7" or "J7" refer to new HCV isolates characterized by the nucleotide sequence 45 disclosed herein, as well as related isolates that are substantially homologous thereto; i.e., at least about 90% or about 95% at the nucleotide level. It is believed that the sequences disclosed herein characterize an HCV subclass that is predominant in Japan and other Asian and/or Pacific rim countries. Additional J1 and J7 isolates can be obtained in view of the disclosure herein and EPO Pub. No. 318,216. In particular, the J1 and J7 nucleotide sequences disclosed herein, as well as the HCV1 sequences in Figure 12, can be used as primers or probes to clone additional domains of J1, J7, 50 or additional isolates.

[0033] As used herein, a nucleotide sequence "from" a designated sequence or source refers to a nucleotide sequence that is homologous (i.e., identical) to or complementary to the designated sequence or source, or a portion thereof. The J1 sequences provided herein are a minimum of about 6 nucleotides, preferably about 8 nucleotides, more preferably about 15 nucleotides, and most preferably 20 nucleotides or longer. The maximum length is the complete 55 viral genome.

[0034] In some aspects of the invention, the sequence of the region from which the polynucleotide is derived is preferably homologous to or complementary to a sequence which is unique to an HCV genome or the J1 and J7 genome. Whether or not a sequence is unique to a genome can be determined by techniques known to those of skill in the art.

For example, the sequence can be compared to sequences in databanks, e.g., Genebank, to determine whether it is present in the uninfected host or other organisms. The sequence can also be compared to the known sequences of other viral agents, including those which are known to induce hepatitis, e.g., HAV, HBV, and HDV, and to other members of the Flaviviridae. The correspondence or non-correspondence of the derived sequence to other sequences can also be determined by hybridization under the appropriate stringency conditions. Hybridization techniques for determining the complementarity of nucleic acid sequences are known in the art. See also, for example, Maniatis et al. (1982) MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.). In addition, mismatches of duplex polynucleotides formed by hybridization can be determined by known techniques, including for example, digestion with a nuclease such as S1 that specifically digests single-stranded areas in duplex polynucleotides. Regions from which typical DNA sequences may be derived include, but are not limited to, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

[0035] The J1 of J7 polynucleotide is not necessarily physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription. In addition, combinations of regions corresponding to that of the designated sequence may be modified in ways known in the art to be consistent with an intended use. The polynucleotides may also include one or more labels, which are known to those of skill in the art.

[0036] An amino acid sequence "from" a designated polypeptide or source of polypeptides means that the amino acid sequence is homologous (i.e., identical) to the sequence of the designated polypeptide, or a portion thereof. An amino acid sequence "from" a designated nucleic acid sequence refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof. The J1 or J7 amino acid sequences in the polypeptides of the present invention are at least about 5 amino acids in length, preferably at least about 10 amino acids, more preferably at least about 15 amino acids, and most preferably at least about 20 amino acids.

[0037] The polypeptides of the present invention are not necessarily translated from a designated nucleic acid sequence; the polypeptides may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from virus. The polypeptides may include one or more analogs of amino acids or unnatural amino acids. Methods of inserting analogs of amino acids into a sequence are known in the art. The polypeptides may also include one or more labels, which are known to those of skill in the art.

[0038] The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is linked to a polynucleotide other than that to which it is linked in nature, or (2) does not occur in nature.

[0039] The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA, and RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example proteins (including for e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

[0040] "Purified polynucleotide" refers to a composition comprising a specified polynucleotide that is substantially free of other components, such composition typically comprising at least about 70% of the specified polynucleotide, more typically at least about 80%, 90% or even 95% to 99% of the specified polynucleotide.

[0041] "Purified polypeptide" refers to a composition comprising a specified polypeptide that is substantially free of other components, such composition typically comprising at least about 70% of the specified polypeptide, more typically at least about 80%, 90% or even 95% to 99% of the specified polypeptide.

[0042] "Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denote microorganisms or higher eukaryotic cell lines cultured as unicellular entities that can be, or have been, used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transformed. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

[0043] A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc. that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

[0044] A "cloning vector" is a replicon that can transform a selected host cell and in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment. Typically, cloning vectors include plasmids, virus (e.g., bacteriophage vector) and cosmids.

[0045] An "integrating vector" is a vector that does not behave as a replicon in a selected host cell, but has the ability

to integrate into a replicon (typically a chromosome) resident in the selected host to stably transform the host.

[0046] An "expression vector" is a construct that can transform a selected host cell and provides for expression of a heterologous coding sequence in the selected host. Expression vectors can be either a cloning vector or an integrating vector.

5 [0047] A "coding sequence" is a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, and recombinant polynucleotide sequences.

[0048] "Control sequence" refers to polynucleotide regulatory sequences which are necessary to effect the expres-
10 sion of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism. In prokaryotes, control sequences generally include promoter, ribosomal binding site, and terminators. In eukaryotes generally control sequences include promoters, terminators and, in some instances, enhancers. The term "control sequences" is intended to include, at a minimum, all components the presence of which are necessary for expression, and may also include additional advantageous components.

15 [0049] "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

[0050] An "open reading frame" or ORF is a region of a polynucleotide sequence which encodes a polypeptide; this
20 region may represent a portion of a coding sequence or a total coding sequence.

[0051] "Immunologically cross-reactive" refers to two or more epitopes or polypeptides that are bound by the same antibody. Cross-reactivity can be determined by any of a number of immunoassay techniques, such as a competition assay.

[0052] As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which comprise at least
25 one epitope. An "antigen binding site" is formed from the folding of the variable domains of an antibody molecule(s) to form three-dimensional binding sites with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows specific binding to form an antibody-antigen complex. An antigen binding site may be formed from a heavy- and/or light-chain domain (VH and VL, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody" includes, without limitation, chimeric antibodies, altered anti-
30 bodies, univalent antibodies, Fab proteins, and single-domain antibodies. In many cases, the binding phenomena of antibodies to antigens is equivalent to other ligand/anti-ligand binding.

[0053] As used herein, a "single domain antibody" (dAb) is an antibody which is comprised of an HL domain, which binds specifically with a designated antigen. A dAb does not contain a VL domain, but may contain other antigen binding domains known to exist to antibodies, for example, the kappa and lambda domains. Methods for preparing dAbs are
35 known in the art. See, for example, Ward et al, Nature 341: 544 (1989).

[0054] Antibodies may also be comprised of VH and VL domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for their preparation are known in the art (see, e.g., U.S. Patent No. 4,816,467, which is incorporated herein by reference), and include the following. For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and heavy chains which are usually
40 aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of the chains are homologous with those sequences found in antibodies produced in vertebrates, whether in situ or in vitro (for example, in hybridomas). Vertebrate antibodies include, for example, purified polyclonal antibodies and monoclonal antibodies, methods for the preparation of which are described infra.

[0055] "Hybrid antibodies" are antibodies where chains are separately homologous with reference to mammalian anti-
45 body chains and represent novel assemblies of them, so that two different antigens are precipitable by the tetramer or aggregate. In hybrid antibodies, one pair of heavy and light chains are homologous to those found in an antibody raised against a first antigen, while a second pair of chains are homologous to those found in an antibody raised against a second antibody. This results in the property of "divalence", i.e., the ability to bind two antigens simultaneously. Such hybrids may also be formed using chimeric chains, as set forth below.

50 [0056] "Chimeric antibodies" refers to antibodies in which the heavy and/or light chains are fusion proteins. Typically, one portion of the amino acid sequences of the chain is homologous to corresponding sequences in an antibody derived from a particular species or a particular class, while the remaining segment of the chain is homologous to the sequences derived from another species and/or class. Usually, the variable region of both light and heavy chains mim-
55 ics the variable regions of antibodies derived from one species of vertebrates, while the constant portions are homologous to the sequences in the antibodies derived from another species of vertebrates. However, the definition is not limited to this particular example. Also included is any antibody in which either or both of the heavy or light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these sources be from differing classes or different species of origin, and whether or not the fusion point is at the variable/con-

stant boundary. Thus, it is possible to produce antibodies in which neither the constant nor the variable region mimic known antibody sequences. It then becomes possible, for example, to construct antibodies whose variable region has a higher specific affinity for a particular antigen, or whose constant region can elicit enhanced complement fixation, or to make other improvements in properties possessed by a particular constant region.

5 [0057] Another example is "altered antibodies", which refers to antibodies in which the naturally occurring amino acid sequence in a vertebrate antibody has been varied. Utilizing recombinant DNA techniques, antibodies can be redesigned to obtain desired characteristics. The possible variations are many, and range from the changing of one or more amino acids to the complete redesign of a region, for example, the constant region. Changes in the constant region, in general, to attain desired cellular process characteristics, e.g., changes in complement fixation, interaction with membranes, and other effector functions. Changes in the variable region may be made to alter antigen binding characteristics. The antibody may also be engineered to aid the specific delivery of a molecule or substance to a specific cell or tissue site. The desired alterations may be made by known techniques in molecular biology, e.g., recombinant techniques, site-directed mutagenesis, etc.

10 [0058] Yet another example are "univalent antibodies", which are aggregates comprised of a heavy-chain/light-chain dimer bound to the Fc (i.e., stem) region of a second heavy chain. This type of antibody escapes antigenic modulation. See, e.g., Glennie et al. *Nature* 295: 712 (1982). Included also within the definition of antibodies are "Fab" fragments of antibodies. The "Fab" region refers to those portions of the heavy and light chains which are roughly equivalent, or analogous, to the sequences which comprise the branch portion of the heavy and light chains, and which have been shown to exhibit immunological binding to a specified antigen, but which lack the effector Fc portion. "Fab" includes aggregates 20 of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as F(ab)₂), which are capable of selectively reacting with a designated antigen or antigen family. Fab antibodies may be divided into subsets analogous to those described above, i.e., "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab". Methods of producing Fab fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

25 [0059] "Epitope" refers to an antibody binding site usually defined by a polypeptide, but also by non-amino acid haptens. An epitope could comprise 3 amino acids in a spatial conformation which is unique to the epitope, generally an epitope consists of at least 5 such amino acids, and more usually, consists of at least 8-10 such amino acids.

[0060] "Antigen-antibody complex" refers to the complex formed by an antibody that is specifically bound to an epitope on an antigen.

30 [0061] "Immunogenic polypeptide" refers to a polypeptide that elicits a cellular and/or humoral immune response in a mammal, whether alone or linked to a carrier, in the presence or absence of an adjuvant.

[0062] "Polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the molecule. Thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

35 [0063] "Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

[0064] A "transformed" host cell refers to both the immediate cell that has undergone transformation and its progeny that maintain the originally exogenous polynucleotide.

[0065] "Treatment" as used herein refers to prophylaxis and/or therapy.

45 [0066] "Individual", refers to vertebrates, particularly members of the mammalian species, and includes but is not limited to domestic animals, sports animals, and primates, including humans.

[0067] "Sense strand" refers to the strand of a double-stranded DNA molecule that is homologous to a mRNA transcript thereof. The "anti-sense strand" contains a sequence which is complementary to that of the "sense strand".

50 [0068] "Antibody-containing body component" refers to a component of an individual's body which is a source of the antibodies of interest. Antibody-containing body components are known in the art, and include but are not limited to, whole blood and components thereof, plasma, serum, spinal fluid, lymph fluid, the external sections of the respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, white blood cells, and myelomas.

[0069] "Purified HCV isolate" refers to a preparation of HCV particles which has been isolated from the cellular constituents with which the virus is normally associated, and from other types of viruses which may be present in the infected tissue. The techniques for isolating viruses are known to those of skill in the art, and include, for example, centrifugation and affinity chromatography.

55 [0070] An HCV "particle" is an entire virion, as well as particles which are intermediates in virion formation. HCV particles generally have one or more HCV proteins associated with the HCV nucleic acid.

[0071] "Probe" refers to a polynucleotide which forms a hybrid structure with a sequence in a target polynucleotide, due to complementarity of at least one region in the probe with a region in the target.

[0072] "Biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, whole blood and components thereof, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of *in vitro* cell culture constituents (including but not limited to conditioned medium resulting from the growth of cells in cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

[0073] The invention pertains to the isolation and characterization of a newly discovered isolate of HCV, J1 and J7, their nucleotide sequences, their protein sequences and resulting polynucleotides, polypeptides and antibodies derived therefrom. Isolates J1 and J7 are novel in their nucleotide and amino acid sequences, and is believed to characteristic of HCV isolates from Japan and other Asian countries.

[0074] The nucleotide sequences derived from HCV/J1 and HCV/J7 are useful as probes to diagnose the presence of virus in samples, and to isolate other naturally occurring variants of the virus. These nucleotide sequences also make available polypeptide sequences of HCV antigens encoded within the J1 and J7 genome and permits the production of polypeptides which are useful as standards or reagents in diagnostic tests and/or as components of vaccines. Antibodies, both polyclonal and monoclonal, directed against HCV epitopes contained within these polypeptide sequences are also useful for diagnostic tests, as therapeutic agents, for screening of antiviral agents, and for the isolation of the NANBH virus. In addition, by utilizing probes derived from the sequences disclosed herein it is possible to isolate and sequence other portions of the J1 and J7 genome, thus giving rise to additional probes and polypeptides which are useful in the diagnosis and/or treatment, both prophylactic and therapeutic, of NANBH.

[0075] The availability of the HCV/J1 and HCV/J7 nucleotide sequences enable the construction of polynucleotide probes and polypeptides useful in diagnosing NANBH due to HCV infection and in screening blood donors as well as donated blood and blood products for infection. For example, from the sequences it is possible to synthesize DNA oligomers of about 8-10 nucleotides, or larger, which are useful as hybridization probes to detect the presence of HCV RNA in, for example, sera of subjects suspected of harboring the virus, or for screening donated blood for the presence of the virus. The HCV/J1 and HCV/J7 sequences also allow the design and production of HCV specific polypeptides which are useful as diagnostic reagents for the presence of antibodies raised during NANBH. Antibodies to purified polypeptides derived from the HCV/J1 and HCV/J7 sequences may also be used to detect viral antigens in infected individuals and in blood.

[0076] Knowledge of these HCV/J1 and HCV/J7 sequences also enable the design and production of polypeptides which may be used as vaccines against HCV and also for the production of antibodies, which in turn may be used for protection against the disease, and/or for therapy of HCV infected individuals. Moreover, the disclosed HCV/J1 and HCV/J7 sequences enable further characterization of the HCV genome. Polynucleotide probes derived from these sequences, as well as from the HCV genome, may be used to screen cDNA libraries for additional viral cDNA sequences, which, in turn, may be used to obtain additional overlapping sequences. See, e.g., EPO Pub. No. 318,216.

[0077] The HCV/J1 and HCV/J7 polynucleotide sequences, the polypeptides derived therefrom and the antibodies directed against these polypeptides, are useful in the isolation and identification of the BB-NANBV agent(s). For example, antibodies directed against HCV epitopes contained in polypeptides derived from the HCV/J1 sequences may be used in processes based upon affinity chromatography to isolate the virus. Alternatively, the antibodies may be used to identify viral particles isolated by other techniques. The viral antigens and the genomic material within the isolated viral particles may then be further characterized.

[0078] The information obtained from further sequencing of the HCV/J1 and HCV/J7 genome, as well as from further characterization of the HCV/J1 and HCV/J7 antigens and characterization of the genomes enable the design and synthesis of additional probes and polypeptides and antibodies which may be used for diagnosis, for prevention, and for therapy of HCV induced NANBH, and for screening for infected blood and blood-related products.

[0079] The availability of HCV/J1 and HCV/J7 cDNA sequences permits the construction of expression vectors encoding antigenically active regions of the polypeptide encoded in either strand. These antigenically active regions may be derived from coat or envelope antigens or from core antigens, or from antigens which are non-structural including, for example, polynucleotide binding proteins, polynucleotide polymerase(s), and other viral proteins required for the replication and/or assembly of the virus particle. Fragments encoding the desired polypeptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta-galactosidase or superoxide dismutase (SOD). Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in EPO Pub. No. 196,056. Vectors encoding fusion polypeptides of SOD and HCV polypeptides are described in EPO Pub. No. 318,216. Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant polypeptide, such as a mature or fusion protein. Alternatively, a polypeptide encoded in the cDNA can be provided by chemical synthesis.

[0080] The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a

signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell is given below. The polypeptide produced in such host cells is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

[0081] Such recombinant or synthetic HCV polypeptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy. In addition, antibodies to these polypeptides are useful for isolating and identifying HCV particles.

[0082] The HCV antigens may also be isolated from HCV virions. The virions may be grown in HCV infected cells in tissue culture, or in an infected host.

[0083] While the polypeptides of the present invention may comprise a substantially complete viral domain, in many applications all that is required is that the polypeptide comprise an antigenic or immunogenic region of the virus. An antigenic region of a polypeptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to regions of HCV/J1 or HCV/J8 epitopes. Accordingly, using the cDNAs of HCV/J1 and HCV/J7 as a basis, DNAs encoding short segments of HCV/J1 and HCV/J7 polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis.

[0084] In instances wherein the synthesized polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier. A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridyl-thio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, *Immun. Rev.* (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus "binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

[0085] Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

[0086] In addition to full-length viral proteins, polypeptides comprising truncated HCV amino acid sequences encoding at least one viral epitope are useful immunological reagents. For example, polypeptides comprising such truncated sequences can be used as reagents in an immunoassay. These polypeptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While these truncated sequences can be produced by various known treatments of native viral protein, it is generally preferred to make synthetic or recombinant polypeptides comprising an HCV sequence. Polypeptides comprising these truncated HCV sequences can be made up entirely of HCV sequences (one or more epitopes, either contiguous or noncontiguous), or HCV sequences and heterologous sequences in a fusion protein. Useful heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the HCV epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, e.g., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783, the disclosures of which are incorporated herein by reference.

[0087] The size of polypeptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, while the maximum size is not critical. In some applications, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated HCV amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select HCV sequences of at

least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

[0088] Truncated HCV amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire viral protein sequence can be screened by preparing a series of short peptides that together span the entire protein sequence. By starting with, for example, 100-mer polypeptides, it would be routine to test each polypeptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified 100-mer to map the epitope of interest. Screening such peptides in an immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare oligopeptides comprising the identified regions for screening. It is appreciated by those of skill in the art that such computer analysis of antigenicity does not always identify an epitope that actually exists, and can also incorrectly identify a region of the protein as containing an epitope.

[0089] The observed relationship of the putative polyproteins of HCV and the Flaviviruses allows a prediction of the putative domains of the HCV "non-structural" (NS) proteins. The locations of the individual NS proteins in the putative Flavivirus precursor polyprotein are fairly well-known. Moreover, these also coincide with observed gross fluctuations in the hydrophobicity profile of the polyprotein. It is established that NS5 of Flaviviruses encodes the virion polymerase, and that NS1 corresponds with a complement fixation antigen which has been shown to be an effective vaccine in animals. Recently, it has been shown that a flaviviral protease function resides in NS3. Due to the observed similarities between HCV and the Flaviviruses, deductions concerning the approximate locations of the corresponding protein domains and functions in the HCV polyprotein are possible. Figure 11 is a schematic of putative domains of the HCV polyprotein. The expression of polypeptides containing these domains in a variety of recombinant host cells, including, for example, bacteria, yeast, insect, and vertebrate cells, should give rise to important immunological reagents which can be used for diagnosis, detection, and vaccines.

[0090] Although the non-structural protein region of the putative polyproteins of the HCV isolate described herein and of Flaviviruses appears to be generally similar, there is less similarity between the putative structural regions which are towards the N-terminus. In this region, there is a greater divergence in sequence, and in addition, the hydrophobic profile of the two regions show less similarity. This "divergence" begins in the N-terminal region of the putative NS1 domain in HCV, and extends to the presumed N-terminus. Nevertheless, it is still possible to predict the approximate locations of the putative nucleocapsid (N-terminal basic domain) and E (generally hydrophobic) domains within the HCV polyprotein. From these predictions it may be possible to identify approximate regions of the HCV polyprotein that could correspond with useful immunological reagents. For example, the E and NS1 proteins of Flaviviruses are known to have efficacy as protective vaccines. These regions, as well as some which are shown to be antigenic in the HCV1, for example those within putative NS3, C, and NS5, etc., should also provide diagnostic reagents.

[0091] The immunogenicity of the HCV sequences may also be enhanced by preparing the sequences fused to or assembled with particle-forming proteins such as, for example, hepatitis B surface antigen or rotavirus VP6 antigen. Constructs wherein the HCV epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and particle-form protein. See, e.g., U.S. Pat. No. 4,722,840; EPO Pub No. 175,261; EPO Pub. No. 259,149; Michelle et al. (1984) Int. Symposium on Viral Hepatitis.

[0092] Vaccines may be prepared from one or more immunogenic polypeptides derived from HCV/J1 or HCV/J7. The observed homology between HCV and Flaviviruses provides information concerning the polypeptides which are likely to be most effective as vaccines, as well as the regions of the genome in which they are encoded. The general structure of the Flavivirus genome is discussed in Rice et al. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, Vol eds. Schlesinger and Schlesinger, Plenum Press). The flavivirus genomic RNA is believed to be the only virus-specific mRNA species, and it is translated into the three viral structural proteins, i.e., C, M, and E, as well as two large nonstructural proteins, NV4 and NV5, and a complex set of smaller non-structural proteins. It is known that major neutralizing epitopes for Flaviviruses reside in the E (envelope) protein. Roehrig (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, Vol eds. Schlesinger and Schlesinger, Plenum Press). The corresponding HCV E gene and polypeptide encoding region may be predicted, based upon the homology to Flaviviruses. Thus, vaccines may be comprised of recombinant polypeptides containing epitopes of HCV E. These polypeptides may be expressed in bacteria, yeast, or mammalian cells, or alternatively may be isolated from viral preparations. It is also anticipated that the other structural proteins may also contain epitopes which give rise to protective anti-HCV antibodies. Thus, polypeptides containing the epitopes of E, C, and M may also be used, whether singly or in combination, in HCV vaccines.

[0093] In addition to the above, it has been shown that immunization with NS1 (nonstructural protein 1), results in protection against yellow fever. Schlesinger et al (1986) J. Virol. 60:1153. This is true even though the immunization does not give rise to neutralizing antibodies. Thus, particularly since this protein appears to be highly conserved among Flaviviruses, it is likely that HCV NS1 will also be protective against HCV infection. Moreover, it also shows that nonstruc-

tural proteins may provide protection against viral pathogenicity, even if they do not cause the production of neutralizing antibodies.

[0094] In view of the above, multivalent vaccines against HCV may be comprised of one or more epitopes from one or more structural proteins, and/or one or more epitopes from one or more nonstructural proteins. These vaccines may be comprised of, for example, recombinant HCV polypeptides and/or polypeptides isolated from the virions. In particular, vaccines are contemplated comprising one or more of the following HCV proteins, or subunit antigens derived therefrom: E, NS1, C, NS2, NS3, NS4 and NS5. Particularly preferred are vaccines comprising E and/or NS1, or subunits thereof. In addition, it may be possible to use inactivated HCV in vaccines; inactivation may be by the preparation of viral lysates, or by other means known in the art to cause inactivation of Flaviviruses, for example, treatment with organic solvents or detergents, or treatment with formalin. Moreover, vaccines may also be prepared from attenuated HCV strains or from hybrid viruses such as vaccinia vectors known in the art [Brown et al. Nature 319: 549-550 (1986)].

[0095] The preparation of vaccines which contain immunogenic polypeptide(s) as active ingredients is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycerol-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

[0096] The vaccines are conventionally administered parenterally, by injection, usually, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

[0097] The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

[0098] The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

[0099] The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and/or reinforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

[0100] In addition, the vaccine containing the immunogenic HCV antigen(s) may be administered in conjunction with other immunoregulatory agents, for example, immune globulins.

[0101] The immunogenic polypeptides prepared as described above are used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker, eds. (1987) IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London).

[0102] Monoclonal antibodies directed against HCV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980) HYBRIDOMA TECHNIQUES; Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS; Kennett et al. (1980) MONOCLONAL ANTIBODIES; see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

[0103] Antibodies, both monoclonal and polyclonal, which are directed against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotypic antibodies.

[0104] Anti-idiotypic antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. Techniques for raising anti-idiotypic antibodies are known in the art. See, e.g., Grzych (1985), *Nature* 316:74; MacNamara et al. (1984), *Science* 226:1325, Uytdehaag et al (1985), *J. Immunol.* 134:1225. These anti-idiotypic antibodies may also be useful for treatment and/or diagnosis of NANBH, as well as for an elucidation of the immunogenic regions of HCV antigens.

[0105] Using the HCV/J1 or HCV/J7 polynucleotide sequences as a basis, oligomers of approximately 8 nucleotides or more can be prepared, either by excision or synthetically, which hybridize with the HCV genome and are useful in identification of the viral agent(s), further characterization of the viral genome(s), as well as in detection of the virus(es) in diseased individuals. The probes for HCV polynucleotides (natural or derived) are a length which allows the detection of unique viral sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of about 10-12 nucleotides are preferred, and about 20 nucleotides appears optimal. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. Among useful probes, for example, are the clones disclosed herein, as well as the various oligomers useful in probing cDNA libraries, set forth below. A complement to any unique portion of the HCV genome will be satisfactory. For use as probes, complete complementarity is desirable, though it may be unnecessary as the length of the fragment is increased.

[0106] For use of such probes as diagnostics, the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. The probes are then labeled. Suitable labels, and methods for labeling probes are known in the art, and include, for example, radioactive labels incorporated by nick translation or kinasin, biotin, fluorescent probes, and chemiluminescent probes. The nucleic acids extracted from the sample are then treated with the labeled probe under hybridization conditions of suitable stringencies. Usually high stringency conditions are desirable in order to prevent false positives. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time, and concentration of formamide. These factors are outlined in, for example, Maniatis, T. (1982) MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

[0107] Generally, it is expected that the HCV genome sequences will be present in serum of infected individuals at relatively low levels, i.e., at approximately 10^2 - 10^3 chimp infectious doses (CID) per ml. This level may require that amplification techniques be used in hybridization assays. Such techniques are known in the art. For example, the Enzo Biochemical Corporation "Bio-Bridge" system uses terminal deoxynucleotide transferase to add unmodified 3'-poly-dT-tails to a DNA probe. The poly dT-tailed probe is hybridized to the target nucleotide sequence, and then to a biotin-modified poly-A. PCT App. No. 84/03520 and EPO Pub. No. 124,221 describe a DNA hybridization assay in which: (1) analyte is annealed to a single-stranded DNA probe that is complementary to an enzyme-labeled oligonucleotide; and (2) the resulting tailed duplex is hybridized to an enzyme-labeled oligonucleotide. EPO Pub. No. 204,510 describes a DNA hybridization assay in which analyte DNA is contacted with a probe that has a tail, such as a poly-dT tail, an amplifier strand that has a sequence that hybridizes to the tail of the probe, such as a poly-A sequence, and which is capable of binding a plurality of labeled strands.

[0108] A particularly desirable technique may first involve amplification of the target HCV sequences in sera approximately 10,000-fold, i.e., to approximately 10^6 sequences/ml. This may be accomplished, for example, by the polymerase chain reactions (PCR) technique described which is by Saiki et al. (1986) *Nature* 324:163, Mullis, U.S. Patent No. 4,683,195, and Mullis et al. U.S. Patent No. 4,683,202. The amplified sequence(s) may then be detected using a hybridization assay which is described in co-pending European Publication No. 317-077 and Japanese application No. 63-260347, which are assigned to the herein assignee, and are hereby incorporated herein by reference. These hybridization assays, which should detect sequences at the level of 10^6 /ml, utilize nucleic acid multimers which bind to single-stranded analyte nucleic acid, and which also bind to a multiplicity of single-stranded labeled oligonucleotides. A suitable

ble solution phase sandwich assay which may be used with labeled polynucleotide probes, and the methods for the preparation of probes is described in EPO Pub. No. 225,807 which is hereby incorporated herein by reference.

[0109] The probes can be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled and the ingredients for labeling may be included in the kit in separate containers. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, for example, standards, wash buffers, as well as instructions for conducting the test.

[0110] Both the HCV/J1 or HCV/J7 polypeptides which react immunologically with serum containing HCV antibodies and the antibodies raised against the HCV specific epitopes in these polypeptides are useful in immunoassays to detect presence of HCV antibodies, or the presence of the virus and/or viral antigens, in biological samples. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. An immunoassay for anti-HCV antibody may utilize one viral epitope or several viral epitopes. When multiple epitopes are used, the epitopes may be derived from the same or different viral polypeptides, and may be in separate recombinant or natural polypeptides, or together in the same recombinant polypeptides.

[0111] An immunoassay for viral antigen may use, for example, a monoclonal antibody directed towards a viral epitope, a combination of monoclonal antibodies directed towards epitopes of one viral polypeptide, monoclonal antibodies directed towards epitopes of different viral polypeptides, polyclonal antibodies directed towards the same viral antigen, polyclonal antibodies directed towards different viral antigens or a combination of monoclonal and polyclonal antibodies.

[0112] Immunoassay protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide. The labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known. Examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

[0113] Typically, an immunoassay for anti-HCV antibody will involve selecting and preparing the test sample, such as a biological sample, and then incubating it with an antigenic (i.e., epitope-containing) HCV polypeptide under conditions that allow antigen-antibody complexes to form. Such conditions are well known in the art. In a heterogeneous format, the polypeptide is bound to a solid support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose, in membrane or microtiter well form, polyvinylchloride, in sheets or microtiter wells, polystyrene latex, in beads or microtiter plates, polyvinylidene fluoride, known as Immobilon™, diazotized paper, nylon membranes, activated beads, and Protein A beads. Most preferably, the Dynatech, Immulon™ 1 microtiter plate or the 0.25-inch polystyrene beads, which Spec finished by Precision Plastic Ball, are used in the heterogeneous format. The solid support is typically washed after separating it from the test sample. In a homogeneous format, the test sample is incubated with antigen in solution, under conditions that will precipitate any antigen-antibody complexes that are formed, as is known in the art. The precipitated complexes are then separated from the test sample, for example, by centrifugation. The complexes formed comprising anti-HCV antibody are then detected by any of a number of techniques. Depending on the format, the complexes can be detected with labeled anti-xenogeneic Ig or, if a competitive format is used, by measuring the amount of bound, labeled competing antibody.

[0114] In immunoassays where HCV polypeptides are the analyte, the test sample, typically a biological sample, is incubated with anti-HCV antibodies again under conditions that allow the formation of antigen-antibody complexes. Various formats can be employed, such as a "sandwich" assay where antibody bound to a solid support is incubated with the test sample; washed; incubated with a second, labeled antibody to the analyte; and the support is washed again. Analyte is detected by determining if the second antibody is bound to the support. In a competitive format, which can be either heterogeneous or homogeneous, a test sample is usually incubated with an antibody and a labeled, competing antigen either sequentially or simultaneously. These and other formats are well known in the art.

[0115] The Flavivirus model for HCV allows predictions regarding the likely location of diagnostic epitopes for the virion structural proteins. The C, pre-M, M, and E domains are all likely to contain epitopes of significant potential for detecting viral antigens, and particularly for diagnosis. Similarly, domains of the nonstructural proteins are expected to contain important diagnostic epitopes (e.g., NS5 encoding a putative polymerase; and NS1 encoding a putative complement-binding antigen). Recombinant polypeptides, or viral polypeptides, which include epitopes from these specific domains may be useful for the detection of viral antibodies in infections blood donors and infected patients. In addition, antibodies directed against the E and/or M proteins can be used in immunoassays for the detection of viral antigens in patients with HCV caused NANBH, and in infectious blood donors. Moreover, these antibodies may be extremely useful in detecting acute-phase donors and patients.

[0116] Antigenic regions of the putative polyprotein can be mapped and identified by screening the antigenicity of bacterial expression products of HCV cDNAs which encode portions of the polyprotein. Other antigenic regions of HCV may be detected by expressing the portions of the HCV cDNAs in other expression systems, including yeast systems and cellular systems derived from insects and vertebrates. In addition, studies giving rise to an antigenicity index and hydrophobicity/hydrophilicity profile give rise to information concerning the probability of a region's antigenicity. Efficient

detection systems may include the use of panels of epitopes. The epitopes in the panel may be constructed into one or multiple polypeptides.

[0117] Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing HCV epitopes or antibodies directed against HCV epitopes in suitable containers, along with the remaining reagents and materials required for the conduct of the assay (e.g., wash buffers, detection means like labeled anti-human Ig, labeled anti-HCV, or labeled HCV antigen), as well as a suitable set of assay instructions.

[0118] The HCV/J1 and HCV/J7 nucleotide sequence information described herein may be used to gain further information on the sequence of the HCV genomes, and for identification and isolation of additional HCV isolates related to J1 or J7. This information, in turn, can lead to additional polynucleotide probes, polypeptides derived from the HCV genome, and antibodies directed against HCV epitopes which would be useful for the diagnosis and/or treatment of HCV caused NANBH.

[0119] The HCV/J1 and HCV/J7 nucleotide sequence information herein is useful for the design of probes for the isolation of additional sequences which are derived from as yet undefined regions of the HCV genomes from which the J1 and J7 sequences are derived. For example, labeled probes containing a sequence of approximately 8 or more nucleotides, and preferably 20 or more nucleotides, which are derived from regions close to the 5'-termini or 3'-termini of the family of HCV cDNA sequences disclosed in the examples may be used to isolate overlapping cDNA sequences from HCV cDNA libraries. These sequences which overlap the cDNAs in the above-mentioned clones, but which also contain sequences derived from regions of the genome from which the cDNA in the above mentioned clones are not derived, may then be used to synthesize probes for identification of other overlapping fragments which do not necessarily overlap the cDNAs described below. Methods for constructing cDNA libraries are known in the art. See, e.g. EPO Pub. No. 318,216. It is particularly preferred to prepare libraries from the serum of Japanese and other Asian patients diagnosed as having NANBH demonstrating antibody to HCV1 antigens; these are believed to be the most likely candidates for carriers of HCV/J1, HCV/J7, or related isolates.

[0120] HCV particles may be isolated from the sera from individuals with NANBH or from cell cultures by any of the methods known in the art, including for example, techniques based on size discrimination such as sedimentation or exclusion methods, or techniques based on density such as ultracentrifugation in density gradients, or precipitation with agents such as polyethylene glycol, or chromatography on a variety of materials such as anionic or cationic exchange materials, and materials which bind due to hydrophobicity.

[0121] A preferred method of isolating HCV particles or antigen is by immunoaffinity columns. Techniques for immunoaffinity chromatography are known in the art, including techniques for affixing antibodies to solid supports so that they retain their immunoselective activity. The techniques may be those in which the antibodies are adsorbed to the support (see, for example, Kurstak in ENZYME IMMUNODIAGNOSIS, page 31-37), as well as those in which the antibodies are covalently linked to the support. Generally, the techniques are similar to those used for covalent linking of antigens to a solid support, described above. However, spacer groups may be included in the bifunctional coupling agents so that the antigen binding site of the antibody remains accessible. The antibodies may be monoclonal, or polyclonal, and it may be desirable to purify the antibodies before their use in the immunoassay.

[0122] The general techniques used in extracting the genome from a virus, preparing and probing a cDNA library, sequencing clones, constructing expression vectors, transforming cells, performing immunological assays such as radioimmunoassays and ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. However, as a general guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

[0123] Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, *E. coli* is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al. (1977), Nature 198:1056, the tryptophan (trp) promoter system (Goeddel et al. (1980) Nucleic Acid Res. 8:4057), and the lambda-derived P_L promoter and N gene ribosome binding site (Shimatake et al. (1981) Nature 292:128) and the hybrid *lac* promoter (De Boer et al. (1983) Proc. Natl. Acad. Sci. USA 292:128) derived from sequences of the *trp* and *lac* UV5 promoters. The foregoing systems are particularly compatible with *E. coli*; if desired, other prokaryotic hosts such as strains of *Bacillus* or *Pseudomonas* may be used, with corresponding control sequences.

[0124] Eukaryotic hosts include yeast and mammalian cells in culture systems. *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Klebsiella lactis* and *Pichia pastoris* are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by

conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (Broach et al. (1983) *Math. Enz.* 101:307), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al. (1968) *J. Adv. Enzyme Eng.* 7:149; Holland et al. (1978), *J. Biol. Chem.* 256:1385), including the promoter for 3-phosphoglycerate kinase (Hitzeman (1980), *J. Biol. Chem.* 255:2073). Terminators may also be included, such as those derived from the enolase gene (Holland (1981), *J. Biol. Chem.* 256:1385). Particularly useful control systems are those which comprise the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion is desired, leader sequence from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are not naturally associated in the wild-type organism. These systems are described in detail in EPO Pub. No. 120,551; EPO Pub. No. 116,201; and EPO Pub. No. 164,556 all of which are incorporated herein by reference.

[0125] Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978), *Nature* 273:113), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or sequences which insure integration of the appropriate sequences encoding NANBV epitopes into the host genome.

[0126] The vaccinia virus system can also be used to express foreign DNA in mammalian cells. To express heterologous genes, the foreign DNA is usually inserted into the thymidine kinase gene of the vaccinia virus and then infected cells can be selected. This procedure is known in the art and further information can be found in these references [Mackett et al. *J. Virol.* 49: 857-864 (1984) and Chapter 7 in *DNA Cloning*, Vol. 2, IRL Press].

[0127] In addition, viral antigens can be expressed in insect cells by the Baculovirus system. A general guide to baculovirus expression by Summer and Smith is *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures* (Texas Agricultural Experiment Station Bulletin No. 1555). To incorporate the heterologous gene into the Baculovirus genome the gene is first cloned into a transfer vector containing some Baculovirus sequences. This transfer vector, when it is cotransfected with wild-type virus into insect cells, will recombine with the wild-type virus. Usually, the transfer vector will be engineered so that the heterologous gene will disrupt the wild-type Baculovirus polyhedron gene. This disruption enables easy selection of the recombinant virus since the cells infected with the recombinant virus will appear phenotypically different from the cells infected with the wild-type virus. The purified recombinant virus can be used to infect cells to express the heterologous gene. The foreign protein can be secreted into the medium if a signal peptide is linked in frame to the heterologous gene; otherwise, the protein will be bound in the cell lysates. For further information, see Smith et al. *Mol. & Cell. Biol.* 3:2156-2165 (1983) or Luckow and Summers in *Virology* 17: 31-39 (1989).

[0128] Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972), *Proc. Natl. Acad. Sci. USA* 69:2110; Maniatis et al. (1982), *MOLECULAR CLONING; A LABORATORY MANUAL* (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al. (1978) *Proc. Natl. Acad. Sci. USA* 75: 1929. Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), *Virology* 52:546 or the various known modifications thereof.

[0129] Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in *Methods in Enzymology* (1980) 65:499-560. Sticky ended cleavage fragments may be blunt ended using *E. coli* DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

[0130] Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation. Ligation mixtures are transformed into suitable

ble cloning hosts, such as *E. coli*, and successful transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

[0131] Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984), DNA 3:401. If desired, the synthetic strands may be labeled with ^{32}P by treatment with polynucleotide kinase in the presence of ^{32}P -ATP, using standard conditions for the reaction. DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982), Nucleic Acids Res. 10:6487.

[0132] DNA libraries may be probed using the procedure of Grunstein and Hogness (1975), Proc. Natl. Acad. Sci. USA 73:3961. Briefly, in this procedure, the DNA to be probed is immobilized on nitrocellulose filters, denatured, and prehybridized with a buffer. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depends on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides such as those derived from cDNA or genomic sequences generally employ higher temperatures, e.g., about 40-42°C, and a high percentage, e.g., 50%, formamide. Following prehybridization, 5'- ^{32}P -labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used as the source of the desired DNA.

[0133] An enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

[0134] To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

Examples

1

[0135] This example describes the cloning of the HCV/J1 and HCV/J7 nucleotide sequences.

[0136] Both blood samples which were used as a source of HCV virions were found to be positive in an anti-HCV antibody assay. The HCV isolates from these samples were named HCV/J1 and HCV/J7. The infectivity of the blood sample containing the J1 isolate was confirmed by a prospective study of blood transfusion recipients. Dr. Tohru Katayama from the Department of Surgery at the National Tokyo Chest Hospital collected blood from patients who have contracted post-transfusion non-A, non-B hepatitis. He also collected blood samples from the respective blood donors of these patients. Next, these samples were assayed for antibodies to the C100-3 HCV1 antigen (EPO Pub. No. 318,216), and blood from one of the donors was found to be positive.

[0137] Isolation of the RNA from the blood samples began by pelleting virions in the blood sample by ultracentrifugation [Bradley, D.W., McCaustland, K.A., Cook E.H., Schable, C.A., Ebert, J.W. and Maynard, J.E. (1985) Gastroenterology 88, 773-779]. RNA was then extracted from the pellet by the guanidinium/cesium chloride method [Maniatis T., Fritsch, E.F., and Sambrook J. (1982) "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory, Cold Spring Harbor] and further purified by phenol/chloroform extraction in the presence of urea, [Berk, A.J. Lee, F., Harrison, T., Williams, J. and Sharp, P.A. (1979) Cell 17, 935-944].

[0138] Five pairs of synthetic oligonucleotide primers were designed from the C/E, E, E/NS1, NS3, and NS5 domains of the nucleotide sequence of HCV1 to isolate fragments from the J1 and J7 genome. The first set of primers were to isolate the sequence from the core and some of the envelope domain. The second set of primers were to isolate the sequences in the envelope domain. The third set of primers were to isolate a fragment which overlapped the putative envelope and non-structural one, NS1, domains. The fourth and fifth set of primers were used to isolate fragments from non-structural domains three and five, NS3 and NS5. The sequences for the various primers are shown below:

The sequence of the primers for the C/E region were:

EP 0 939 128 A2

21S 5' CGTGCCCCCGCAAGACTGCT 3'
J80A 5' CCGTCCTCCAGAACCCGGAC 3'

5

The sequence of the primers for the E region were:

71S 5' GCCGACCTCATGGGGTACAT 3'
J132A 5' AACTGCGACACCACTAAGGC 3'

10

15 The sequence of the primers for the E/NS1 region were:

127S 5' TGGCATGGGATATGATGATG 3'
166A 5' TTGAACTTGTGGTGATAGAA 3'

20

The sequence of the primers for the NS3 region were :

464S 5' GGCTATACCGGCGACTTCGA 3'
526A 5' GACATGCATGTCATGATGTA 3'

25

The sequence of the primers for the NS5 region were:

870S 5' GCTGGAAAGAGGGTCTACTA 3'
917A 5' GTTCTTACTGCCCAGTTGAA 3'

30

35

[0139] 1 µg of the antisense primers, 166A, 526A, or 917A, was added to 10 units of reverse transcriptase (Biorad) to synthesize cDNA fragments from the isolated RNA as the template. The cDNA fragments were then amplified by a standard polymerase chain reaction [Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn G.T., Erlich, H.A., and Arnheim, N. (1985) Science 230, 1350-1354] after 1 µg of the appropriate sense primer, 21S, 71S, 127S, 464S or 870S, was added.

40

[0140] The cDNA fragments amplified by the PCR method were gel isolated and cloned by blunt-end ligation into the SmaI site of pUC119 [Vieira, J. and Messing, J. (1987) Methods in Enzymology 153, 3-11] or into the SnaBI site of charomid SB, a derivative of the cloning vector charomid 9-42 [Saito, I. and Stark, G. (1986) Proc. Natl. Acad. Sci. USA 83: 8664-8668]. Clones which contain the fragments of the five viral domains were successfully constructed.

45

II

[0141] From the PCR reaction of the Japanese isolates, J1 and J7, three independent clones from each region, C/E, E, E/NS1, NS3, and NS5, have been sequenced by the dideoxy chain termination method.

50

[0142] Sequence from all regions except C/E has been isolated from the J1 isolate. Sequence from only the C/E region has been isolated from the J7 isolate. Surprisingly, fragments isolated from both isolates are neither longer or shorter than what would be predicted from the HCV1 genome. However, there is heterogeneity between clones containing sequence from the same region. Consequently, a consensus sequence was constructed for each of the domains, C/E, E, E/NS1, NS3 and NS5, as shown respectively in Figures 1 through 5. These differences may be explained as artifacts which occur randomly during the PCR amplification [Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985) Science 230, 1350-1354]. Another explanation is that more than one virus genome is present in the plasma of a single healthy carrier and that these genomes are heterogeneous at the

55

nucleotide level.

[0143] To clarify this point, it was determined how many of these nucleotide differences would lead to amino acid changes, using the sequence from the NS3 domain of the J1 isolate as an example. Out of the five nucleotide differences, three fall on the third position of the amino acid codon and do not change the amino acid sequence. Both of the remaining two nucleotide changes fall on the first position of the amino acid codon and generate amino acid changes of threonine to alanine and proline to alanine, all of which are small, neutral amino acid residues. Similarly, when analyzing the nucleotide differences in other domains, many silent and conserved mutations are found. These results suggest that nucleotide sequences of the HCV genomes in the plasma of a single healthy donor are heterogeneous at the nucleotide level.

[0144] In addition, once the consensus sequences for each of the fragments were compiled each sequence was compared to the HCV1 isolate in Figures 6 through 10. In Figure 6 the fragment from the C/E region of the J7 isolate shows a 92.8%, 512/552, nucleotide and 97.4%, 150/154, amino acid homology to the HCV1 isolate. The fragment from the E domain of J1 shows a slightly lower nucleotide and amino acid homology to HCV1 in Figure 7 of 76.2% and 82.9%, respectively. The fragment from the J1 isolate which overlaps the envelope and non-structural one domains shows the lowest homology to HCV1, as seen in Figure 8, where the J1 isolate has a 71.5% nucleotide homology and a 73.5% amino acid homology to HCV1. Figure 9 shows a comparison of the fragment from the NS3 domain of J1 to HCV1. The homology between the nucleotides sequences is 79.8%, while the amino acid homology between the isolates is quite high, 92.2% or 179/194 amino acids. Figure 10 shows the homology between the NS5 sequences from J1 and HCV1. The sequences have a 84.3% nucleotide and 88.7% amino acid homology.

[0145] The vectors described in the examples above were deposited with the Patent Microorganism Depository, Fermentation Institute, Agency of Industrial Science and Technology at 1-3, Higashi 1-chome Tsukuba-chi, Ibaragiken 305, Japan, and will be maintained under the provisions of the Budapest Treaty. The accession numbers and dates of the deposit are listed below, on page 68.

III

[0146] An HCV/J1 clone, J1-1519, was isolated using the essentially the techniques described above. However, the primers used in the isolation were J159S and 199A. The sequences of the oligomeric primers J159S and 199A, which follow, were based upon those in J1-1216 and in HCV1.

J159S	5' ACT GCC CTG AAC TGC AAT GA 3'
199A	5' AAT CCA GTT GAG TTC ATC CA 3'

35

[0147] Clone J1-1519 is comprised of an HCV cDNA sequence of 367 nucleotides which spans most of the 5'-half of the NS1 region and which overlaps the E-region clone, J1-1216, by 31 nucleotides. Three independent clones spanning this region were sequenced; the sequences in this region obtained from the three clones were identical. The sequence of the HCV cDNA in J1-1216 (shown in the figure as J1) and the amino acids encoded therein (shown above the nucleotide sequence) are shown in Figure 13. Figure 13 also shows the sequence differences between J1-1216 in the comparable region of the prototype HCV1 cDNA (indicated in the figure as PT), and the resulting changes in the encoded amino acids. The homology between the J1-1216 and HCV1 cDNA is approximately 70% at the nucleotide level, and about 75% at the amino acid level.

[0148] A composite of the sequences from the putative core to NS1 region of the J1 isolate is shown in Figure 14; also shown in the figure are the amino acids encoded in the J1 sequence. The variation from the HCV1 prototype sequence is shown in the line below the J1 nucleotide sequence; the dashed lines indicate homologous sequences. The nonhomologous amino acid encoded in the HCV1 prototype sequence is shown below the HCV1 nucleotide sequence.

[0149] Cloned material containing the J1/1519 HCV cDNA (pS1-1519) has been maintained in DH5 α , and deposited with the Patent Microorganism Depository.

IV

[0150] Several regions of the J1 isolate, including the C200-C100 region from the putative NS3-NS4 region (which encompasses the region encoding the 5-1-1 polypeptide in HSV1 (See EPO Pub. No. 318,216), and the putative NS1-E region, were amplified using the PCR method. The C200-C100 region includes nucleotides 3799 to 5321 of the prototype HCV1. RNA was extracted as described above, except that extraction was with guanidinium thiocyanate in the

EP 0 939 128 A2

presence of Proteinase K and sodium dodecylsulfate (SDS) (Maniatis (1982), supra). The RNA was transcribed into HCV cDNA by incubation in a 25 µl reaction comprised of 1 µM of each primer, 40 units of RNase inhibitor (RNASIN), 5 units of AMV reverse transcriptase, and salts and buffer necessary for the reaction. Amplification of a segment of the HCV cDNA from the designated region was performed utilizing pairs of synthetic oligomer 16-mer primers. PCR amplification was accomplished in three rounds (PCR I, PCR II, and PCRIII). The second and third rounds of PCR amplification (PCR II) utilized different sets of PCR primers; the first PCR reaction was diluted 10-fold and multiple rounds of PCR amplification were carried out with the new primers, so that ultimately up to 50% of the products of the first PCR reaction (PCR I) were reamplified. The primers used for the amplification of the regions were the following. These primers, with the exception of J1C200-3 which was derived from the J1 isolate sequence, were derived from the prototype HCV1 sequence.

Primers for amplification of the "5-1-1" region from NS3-NS4.

PCR I

[0151]

511/16A (sense, derived from nucleotides starting at number 1528 of HCV1)

5' AAC AGG CTG CGT GGT C 3'

511/16B (anti-sense, derived from nucleotides ending at 5260 of HCV1)

5' AGT TGG TCT GGA CAG C 3'

PCR II

[0152]

511/35A (sense, the HCV portion derived from nucleotides starting at number 5057 of HSV1; the restriction enzyme site is underlined)

5' CTTGAATTC TCG TCT TGT CCG GGA AGC CGG CAA TC 3'

511/35B (anti-sense, the HCV portion derived from nucleotides ending at number 5233 of HSV1; the restriction enzyme site is underlined)

5' CTTGAATTC CCT CTG CCT GAC GGG ACG CGG TCT GC 3'

PCRIII

[0153]

511/35A (see supra)

VSNrc7 (antisense, derived from nucleotides ending at number 5804 of HSV1)

5' GTA GTG CGT GGG GGA AAC AT 3'

Primers for amplification of the "NS1/E" region

PCR I

[0154]

J1(E2)3 (sense, the HCV portion derived from nucleotides starting at number 953 of HSV1, the restriction enzyme site is underlined)

5' CTTAGAATTC TGG CAT GGG ATA TGA TGA TG 3'

J1(E)4 (sense, the HCV portion derived from nucleotides starting at number 1087 of HSV1, the restriction enzyme site is underlined)

5' CTTAGAATTC TCC ATG GTG GGG AAC TGG GC 3'

J1rc12 (anti-sense, the HCV portion derived from nucleotides ending at 1995 of HSV1, the restriction enzyme site is underlined)

5' CTTGAATTC TAA CGG GCT GAG CTC GGA 3'

J1rc13 (anti-sense, the HCV portion derived from nucleotides ending at 1941 of HSV1, the restriction enzyme site is underlined)

EP 0 939 128 A2

5' CTTAGAATTC CGT CCA GTT GCA GGC AGC TTC 3'

PCR II

5 [0155]

J1rc13 (see supra)

J1IZ-1 (sense, the HCV portion is derived from nucleotides starting at number 1641 of HCV1, the restriction enzyme site is underlined)

10 5' CTTGAATTC CAA CTG GTT CGG CTG TAC A 3'

J1IZ-2 (sense, the HCV portion is derived from nucleotides starting at number 1596 of HCV1, the restriction enzyme site is underlined)

5' TGA GAC GGA CGT GCT GCT CCT 3'

15 Primers for the C200-C100 region of the "NS3-NS4" region

PCR I

[0156]

20

J1C200-1 (sense, derived from nucleotides starting at number 3478 of HCV1)

5' TCC TAC TTG AAA GGC TC 3'

J1C200-3 (anti-sense, derived from nucleotides ending at number 4402 of HCV1)

5' GGA TCC AAG CTG AAA TCG AC 3'

25

J1rc52 (anti-sense, the HCV portion derived from nucleotides ending at 5853 of HCV1, the restriction enzyme site is underlined)

5' CTTAGAATTC GAG GCT GCT GAG ATA GGC AGT 3'

511/16A (see above).

30 PCR II

[0157]

35 J1C200-2 (sense, the HCV portion derived from nucleotides starting at number 3557 of HCV1, the restriction enzyme site is underlined)

5' CTTGAATTC CCC GTG GAG TGG CTA AGG CGG TGG ACT 3'

J1C200-4 (anti-sense, the HCV portion derived from nucleotides ending at 4346 of HCV1, the restriction enzyme site is underlined)

5' CTTGAATTC TCG AAG TCG CCG GTA TAG CCG GTC ATG 3'

40

511/35A (see above)

J1rc51 (anti-sense, the HCV portion derived from nucleotides ending at 5826 of HCV1, the restriction enzyme site is underlined)

5' CTTAGAATTC GGC AGC TGC ATC GCT CTC CGG CAC 3'

45

The amplified HCV cDNAs were either sequenced directly without cloning, and/or were cloned. Sequencing was accomplished using an asymmetric PCR technique, essentially as described in Shyamala and Ames, J. Bacteriology 171:1602 (1989). In this technique, amplification of the cDNA is carried out with a limiting concentration of one of the primers (usually in a ratio of about 1:50) in order to get preferential amplification of one strand. The preferentially amplified strand is then sequenced by the dideoxy chain termination method.

50 The primers used for asymmetric sequencing by the PCR method were the following. For the NS1 region: J1IZ-1 and J1rc13 (sequenced with both); J1IZ-2, J1rc13 (confirmed on both strands). For the NS3-NS4 region, which includes the C200-C100 N-terminal region, C200-C100 C-terminal region, and the 5-1-1 region: J1C200-2 and J1C200-7 (for the N-terminal region of C200-C100), and J1C200-4 and J1C200-6 (for the C200-C100 C-terminal region); and 511/35A and hep 4 (for the 5-1-1 region). The sequences for J1C200-2, J1C200-4, and 511/35A are shown supra; the sequences of hep 4, J1C200-6, and J1C200-7 are the following.

55

hep 4 (derived from nucleotides starting at number 5415 of HCV1)

5' TT GGC TAG TGG TTA GTG GGC TGG TGA CAG 3'

J1C200-6 (the HCV portion derived from nucleotides starting at number 3875 of HCV1, the restriction enzyme site is underlined)

5' CTTGAATTC CGT ACT CCA CCT ACG GCA AGT TCC TT 3'

J1C200-7 (the HCV portion derived from nucleotides starting at number 3946 of HCV1, the restriction enzyme site is underlined)

5' CTTGAATTC GTG GCA TCC GTG GAG TGG CAC TCG TC 3'

5

[0158] The sequences obtained by asymmetric sequencing of the "NS1" region, the C200-C100 region, and the 5-1-1 region are shown in Figure 15, and Figure 16, respectively. In the figures, the amino acids encoded in the J1 sequence are shown above the J1 nucleotide sequence. The differences between the J1 sequence and the HCV1 prototype nucleotide sequence is shown below the J1 sequence (the dashes indicate homologous nucleotides in both sequences). The encoded amino acids which differ in the HCV1 prototype sequence are shown below the HCV1 nucleotide sequence.

[0159] HCV cDNAs from the NS1 region, the C200-C100 region, and the 5-1-1 region were cloned. A 300 bp and a 230 bp fragment from the putative NS1 region, were cloned into a derivative of the commercially available vector, pGEM-3Z, in host HB101, and deposited with the ATCC as AW-300bp. The derivative vectors maintain the original pGEM-3Z polylinkers, an intact Amp^r gene, and the genes required for replication in *E. coli*. The HCV cDNA fragments may be removed with SacI and XbaI. HCV cDNAs containing 770 bp N-terminal fragments of C200 were cloned into pM1E in HB101, 12 clones were pooled and deposited with the ATCC as AW-770bp-N; the HCV cDNA may be removed from the vector with HaeII. The resultant HaeII fragment will contain vector DNA of 300 bp and 250 bp at the 5' and 3' ends, respectively. HCV cDNAs containing 700 bp C-terminal fragments of C200 (AW-700bp-C) were cloned into M13mp10 and maintained in host DH5 α -F⁻; cloning was into the vector polylinker site. The resultant phage were pooled, and deposited with the ATCC on September 11, 1990 as AW-700bp-N or AW-700bp-C. HCV cDNA from J1 equivalent to the 5-1-1 region of HCV1 was cloned into mp19 R1 site, and maintained in DH5 α -F⁻. Several m13 phage superanants from this cloning were pooled and deposited with the ATCC as J1 5-1-1, on September 11, 1990. The HCV cDNAs may be obtained from the phage by treatment with EcoRI. Accession numbers for J1 5-1-1 and AW-700bp-N or AW-700bp-C may be obtained by telephoning the ATCC at (301) 881-2600.

25

[0160] The above-described cloned material was deposited with the American Type Culture Collection (ATCC).

V

[0161] An HCV cDNA library containing sequences of the putative "NS1" region of the J1 isolate was created by directional cloning in λ -gt22. The "NS1" region extends from about nucleotide 1460 to about nucleotide 2730 using the numbering system of the HCV1 prototype nucleic acid sequence, where nucleotide 1 is the first nucleotide of the initiating methionine codon for the putative polyprotein. The cloning was accomplished using essentially the method described by Han and Rutter in GENETIC ENGINEERING, Vol 10 (J.K. Setlow, Ed., Plenum Publishing Co., 1988), except that the primers for the synthesis of the first and second strand of HCV cDNA were JHC67 and JHC68, respectively, and the source of RNA was the J1 plasma. In this method the RNA is extracted with guanidium thiocyanate at a low temperature. The RNA is then converted to full length cDNA, which is cloned in a defined orientation relative to the *lacZ* promoter in λ -phage. Using this method, the HCV cDNAs to J1 RNA were inserted into the NotI site of λ -gt22. The presence of "NS1" sequences in the library was detected using as probe, Alx54.

[0162] The sequence of a region of "NS1" downstream from the region shown in Figure 14, but which overlaps the region by about 20 nucleotides, was determined using the asymmetric sequencing technique described above, but substituting as primers for PCR amplification, Alx 61 and Alx 62. The resulting sequence is shown in Figure 17. (It should be noted that the PCR amplification was of a region from about nucleotide 1930 to about nucleotide 2340; this region is also encompassed in the sequence shown in Figure 15). The sequences of the primers and probes used to obtain the HCV cDNA library in λ -gt22, and to sequence the portion of the "NS1" region were the following.

JHC 67
5' GACGC GGCCG CCTCC GTGTC CAGCG CGT 3'
JHC 68
5' CGTGC GGCCG CAAGA CTGCT AGCCG AGGT 3'
ALX 61
5' ACCTG CCACT GTGTA GTGGT CAGCA GTAAC 3'
ALX 62
5' ACGGA CGTCT TCGTC CTTAACAATA CCAGG 3'
ALX 54
5' GAACT TTGCG ATCTG GAAGACAGGG ACAGG 3'

[0163] A 400 bp fragment of J1 HCV cDNA derived from the sequenced region was cloned into pGEM3z and main-

tained in HB101; the HCV cDNA may be removed from the vector with SacI and XbaI. Host cells transformed with the vector (JH-400bp) have been deposited with the ATCC.

[0164] A pooled cDNA library was created from the J1 serum; the pooled library spans the J1 genome and is identified as HCV-J1 λ gt22. The pooled cDNA library was created by pooling aliquots of 11 individual cDNA libraries, which had been prepared using the directional cloning technique described above, except that the libraries were created from primers which were designed to yield HCV cDNAs which spanned the genome. The primers were derived from the sequence of HCV1, and included JHC 67 and JHC 68. The HCV cDNAs were inserted into the NotI site of λ -gt22. The pooled cDNA library, HCV-J1 λ gt22, has been deposited with the ATCC.

10 VI

[0165] The sequence of a region of the polynucleotide upstream of that shown in Figure 14 was determined. This region begins at nucleotide -267 with respect to the HCV1 (See Figure 12) and extends for 560 nucleotides. Sequencing was accomplished by preparing HCV cDNA from RNA extracted from J1 serum, and amplifying the HCV cDNA using the PCR method.

[0166] RNA was extracted from 100 μ l of serum following treatment with proteinase K and sodium dodecylsulfate (SDS). The samples were extracted with phenol-chloroform, and the RNA precipitated with ethanol.

[0167] HCV cDNA from the J1 isolate was prepared by denaturing the precipitated RNA with 0.01M MeHgOH; after ten minutes at room temperature, 2-mercaptoethanol was added to sequester the mercury ions. Immediately, the mix for the first strand of cDNA synthesis was added, and incubation was continued for 1 hr at 37°C. The conditions for the synthesis of the anti-sense strand were the following: 50 mM Tris HCl, pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, 500 μ M each deoxynucleotide triphosphate, 250 pmol specific antisense cDNA primer r25, 250 units MMLV reverse transcriptase. In order to synthesize the second strand (sense), the synthesis reaction components were added, and incubated for one hour at 14°C. The components for the second strand reaction were as follows: 14 mM Tris HCl, pH 8.3, 68 mM KCl, 7.5 mM ammonium sulfate, 3.5 mM MgCl₂, 2.8 mM dithiothreitol, 25 units DNA polymerase I, and one unit RNase H. The reactions were terminated by heating the samples to 95°C for 10 minutes, followed by cooling on ice.

[0168] The HCV cDNA was amplified by two rounds of PCR. The first round was accomplished using 20 μ l of the cDNA mix. The conditions for the PCR reaction were as follows: 10 mM Tris HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.002% gelatin, 200 mM each of the deoxynucleotide triphosphates, and 2.5 units Amplitaq. The PCR thermal cycle was as follows: 94°C one minute, 50°C one minute, 72°C one minute, repeated 40 times followed by seven minutes at 72°C. The second round of PCR was accomplished using nested primers (i.e. primers which bound to an internal region of the first round of PCR amplified product) to increase the specificity of the PCR products. One percent of the first PCR reaction was amplified essentially as the first round, except that the primers were substituted, and the second step in the PCR reaction was at 60°C instead of 50°C. The primers used for the first round of PCR were ALX90 and r14. The primers used for the second round of PCR were r14 and p14.

[0169] The sequences of the primers for the synthesis of HCV cDNA and for the PCR method were the following.

```

r25
40 5' ACC TTA CCC AAA TTG CGC GAC CTA 3'
ALX90
    5' CCA TGA ATC ACT CCC CTG TGA GGA ACT A 3'
r14
    5' GGG CCC CCAG CTA GGC CGA GA 3'
45 p14
    5, AAC TAC TGT CTT CAC GCA GAA AGC 3'

```

[0170] The PCR products were gel purified, the material which migrated as having about 615 bp was isolated, and sequenced by a modification of the Sanger dideoxy chain termination method, using ³²P-ATP as label. In the modified method, the sequence replication was primed using P32 and R31 as primers; the double stranded DNA was melted for 3 minutes at 95°C prior to replication, and the synthesis of labeled dideoxy terminated polynucleotides was catalyzed by Bst polymerase (obtained from BioRad Corp.), according to the manufacturer's directions. The sequencing was performed using 500ng to 1 μ g of PCR product per sequencing reaction.

[0171] The primers P32 (sense) and R31 (antisense) were derived from nucleotides -137 to -115 and from nucleotides 192 to 173, respectively, of the HCV1 sequence. The sequences of the primers are the following.

```

P32 primer
    5' AAC CCG CTC AAT GCC TGG AGA TT 3'

```

EP 0 939 128 A2

R31 primer

5' GGC CGX CGA GCC TTG GGG AT 3'

where X = A or G

5 [0172] The sequence of the region in the J1 isolate which encompasses the 5'-untranslated region as well as a part of the region of the putative "Core" is shown in Figure 18. In the figure, amino acids encoded in the J1 sequence are shown above the nucleotide sequence. The sequence of the prototype HCV1 is shown below the J1 sequence; the dashes indicate sequence homology with J1. The differing amino acids encoded in the HCV1 sequence are shown below the HCV1 sequence.

10 [0173] An HCV cDNA fragment which is a representative of the 600 bp J1 sequence described above (TC 600bp) was cloned into pGEM3Z and maintained in host HB101; the HCV cDNA fragment may be removed with SacI and XbaI. This material is on deposit with the ATCC.

Patent Microorganism Depository-deposited under Budapest Treaty terms.

15 [0174]

Deposited Materials	Accession Number	Deposit Date
<u>E. coli</u> DH5/pS1-8791a (This clone contains 427 bp of the HS5 domain of J1)	BP-2593	9/15/1989
<u>E. coli</u> HB101/pU1-1216c (This clone contains 351 bp of the E/NS1 domains of J1)	BP-2594	9/15/1989
<u>E. coli</u> HB101/pU1-4652d (This clone contains 583 bp of the NS3 domain of J1)	BP-2595	9/15/1989
<u>E. coli</u> DH5α/pS1-713c (This clone contains 580 bp of the E domain of J1)	BP-2637	11/1/1989
<u>E. coli</u> DH5α/pS7-28c (This clone contains 552 bp of the C/E domain of J7)	BP-2638	11/1/1989
<u>E. coli</u> DH5α/ps1-1519	BP3081	8/30/90

40 [0175] The following vectors described in the Examples were deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers. The deposits were made under the terms of the Budapest Treaty.

Deposited Materials	Accession Number	Deposit Date
TC-600BP (in <u>E. coli</u> HB101/pGEM3Z)	68393	9/11/90
JH-400bp (in <u>E. coli</u> HB101/pGEM3Z)	68394	9/11/90
AW-300bp (in <u>E. coli</u> HB101/pGEM3Z)	68392	9/11/90
AW-770bp-N (in <u>E. coli</u> HB101/pM1E)	68395	9/11/90
AW-700bp-C or AW-700bp-N (in <u>E. coli</u> DH5α-F/M13mp10)		
J1 5-1-1 (in <u>E. coli</u> DH5α-F/M13mp10)		
HCV-J1 λ gt22	40884	9/6/90

These deposits are provided for the convenience of those skilled in the art. These deposits are neither an admission

EP 0 939 128 A2

that such deposits are required to practice the present invention nor that equivalent embodiments are not within the skill of the art in view of the present disclosure. The public availability of these deposits is not a grant of a license to make, use or sell the deposited materials under this or any other patent. The nucleic acid sequences of the deposited materials are incorporated in to present disclosure by reference, and are controlling if in conflict with any sequences described herein.

5

[0176] While the present invention has been described by way specific examples for the benefit of those in the field, the scope of the invention is not limited as additional embodiments will be apparent to those of skill in the art from the present disclosure.

10

15

20

25

30

35

40

45

50

55

SEQUENCE LISTING

- 5 (1) GENERAL INFORMATION:
- (i) APPLICANT: CHIRON CORPORATION
- 10 (ii) TITLE OF INVENTION: New HCV Isolates
- (iii) NUMBER OF SEQUENCES: 59
- (iv) CORRESPONDENCE ADDRESS:
- 15 (A) ADDRESSEE: J. A. Kemp & Co.
 (B) STREET: 14 South Square Gray's Inn
 (C) CITY: London
 (D) STATE:
 (E) COUNTRY: United Kingdom
 (F) ZIP: WC1R 5LX
- 20 (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 25 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER: EP 90310149.1
 (B) FILING DATE: 17-SEP-1990
 30 (C) CLASSIFICATION: C12N 15/51
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER:
 (B) FILING DATE:
- 35 (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Goldin, Douglas M.
 (C) REFERENCE/DOCKET NUMBER: N.61241-DMG/kst
- 40 (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: 071-405-3292
 (B) TELEFAX: 071-242-8932
 (C) TELEX: 23676
- 45 (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 50 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
- 55

CGTGCCCCCG CAAGACTGCT

5 (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 10 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

15 CCGTCTCTCCA GAACCCGGAC

20

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGCGAAGCTCA TGGGGTACAT

20

30 (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

40 AACTGGGACA CCACTAAGGC

20

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 45 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

55

TGGCATGGGA TATGATGATG

20

5

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

15

TTGAACTTGT GGTGATAGAA

20

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

25

GGCTATACCG GCGACTTCGA

20

30

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

40

GACATGCATG TCATGATGTA

20

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

50

55

GCTGGAAAGA GGGTCTACTA

20

5

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

15

GTTCTTACTG OCCAGTTGAA

20

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

20

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACTGCCCTGA ACTGCAATGA

20

30

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

40

AACCAGTTGA GTTCATCCA

19

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

45

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

55

AACAGGCTGC GTGGTC

16

5 (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 16 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

15 AGTGGGCTCTG GACAGC

16

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CITGAATTCT CGTCTGTGTC GGAAGCGG CAATC

35

30 (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

40 CITGAATTCC CTCGCGCTGA CGGAGCGGG TCTGC

35

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

55

GTAGTGCGTG GGGGAAACAT 20

5 (2) INFORMATION FOR SEQ ID NO:18:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

15 CTTAGAATTC TGGCATGGGA TATGATGATG 30

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CTTAGAATTC TOCATGGTGG GGAAC TGGGC 30

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

40 CTTAGAATTC CGTCCAGTTC CAGGCAGCTT C 31

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

50 CTTGAATTCC AACTGGTTTC GCTGTAC 27

55

(2) INFORMATION FOR SEQ ID NO:22:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGAGACGGAC GTGCTGCTCC T 21

15 (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

25 TCCTACTTGA AAGGCTC 17

(2) INFORMATION FOR SEQ ID NO:24:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGATCCAAGC TGAAATGAC 20

40 (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 45 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

50 CTTAGAATTC GAGGCTGCTG AGATAGGCAG T 31

55

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTTGAATTCC CGGTGGAGTG GCTAAGGCGG TGGACT

36

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CTTGAATTCT CGAAGTGGCC GGTATAGCGG GTCATG

36

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTTAGAATTG GGCAGCTGCA TCGCTCTCGG GCAC

34

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TTGGCTAGTG GTTAGTGGGC TGGTGACAG

29

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTTGAATTCC GTACTCCACC TAAGGCAAGT TCCTT

35

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CTTGAATTCC TGGCATCCGT GGAGTGGCAC TGTC

35

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GACGGGGCCG CCTCCGTGTC CAGCGCGT

28

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CGTGGGGCCG CAAGACTGCT AGCGAGGT

29

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ACCTGOCAC TGTAGTGGT CAGCAGTAAC

30

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

ACGGACGTCT TGTCCCTTAA CAATACCAGG

30

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GAACITTTGCG ATCTGGAAGA CAGGGACAGG

30

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

ACCTTACCCA AATTGCGCGA CCTA

24

(2) INFORMATION FOR SEQ ID NO:38:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

10 CCATGAATCA CTCOCCTGTG AGGAACTA 28

(2) INFORMATION FOR SEQ ID NO:39:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGGCCCCCAG CTAGGCGAG A 21

25 (2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 30 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

35 AACTACTGTC TTCAAGCAGA AAGC 24

(2) INFORMATION FOR SEQ ID NO:41:

- 40 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AACCCGCTCA ATGCTGGAG ATT 23

50 (2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GGCCGCGAG OCTTGGGGAT

20

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 552 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AGCCGAGTAG TGTTGGGTGG CGAAAGGCCT TGTTGGTACTG CCTGATAGGG TGCTTGCGAG 60
 TGCCCCGGGA GGTCTGGTAG ACGTGCATC ATGAGCACAA ATCTAAACC YCAAAGAAAA 120
 ACCAAACGTA ACACCAACCG TGGCCACAG GACGTYAAGT TCCCKGGGG TGCTCAGATC 180
 GTYGGTGGAG TTTACTTGTT GCGCGCAGG GGCCCCAGGT TGGGTGTGG TGCGACTAGG 240
 AAGACTTCGG AGCGGTCRCA ACCTGTGGA AGGCGACAAC CTATCCCCAA GGCTGGCGG 300
 CCGAGGGCA GGACCTGGGC TCAGCCTGGG TATCCTTGGC CCTCTATGG CAATGAGGGC 360
 TWGGGGTGGG CAGGATGGCT CCTGTCAACC CGCGCTCTC GGCTAGTTG GGGCCYAMT 420
 GACCCCGGC GTAGGTGGG TAATTTGGGT AAGGTCATG ATACCTTAC ATGCGGCYTC 480
 GCGACCTCA TGGGGTACAT YCGCTYGTG GCGCCCCCT TAGGGGGGCG TGCCAGGGCC 540
 CTGGCACATG GT 552

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 580 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TCCGCTGTC GCGCCCCCT YAGGGGGGCG TGCCAGGGCC CTGGCACATG GTGTCCGGT 60

EP 0 939 128 A2

TCTGGAGGAC GGGTGAACT ATGCAACAGG GAATTGCCC GGTTGCTCTT TCTCTATCTT 120
 CCTCTTTCCT YTGCTGTCTT GTTTGAACAT CCCAGCTTCC GCTTATGAAG TGGGCAAGT 180
 GTCGGGGATA TAYCATGTCA CAAACGACTG CTCCAACTCA AGCAATTGIGT ATGAGGGGGC 240
 GGAOFTGATC ATGCATGCCO COGGGTGGGT GOCCTGGGTT OGGGAGAACA AYTCCTCCCG 300
 TTGCTGGGTA GGGCTCACTC CCAOGCTGCG GGCCAGGAAT GCCAGGTTCC CCACTACGAC 360
 ATTROGAGCG CAOGTCGACT TGCTGGTTGG GAOGGCTGCT TTCTGCTCCG CTATGTACGT 420
 GGGGGATCTC TGGGATCTG TTTTCTYAT CTCACAGCTG TTCACCTTCT GGGCTGGCG 480
 GCATGAGACA GTACAGGACT GCAACTGCTC AATCTATCCC GGCCAAGTAT CAGGCCATCG 540
 YATGGCTTGG GATATGATGA TGAACGTGTC GCCCACGGCA 580

20

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 351 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

30

AACTGGTGGC CCAOGGCAGC CTTAGTGGTG TGCAGTTAC TCGGATCCC ACAAGCTGTC 60
 ATGGACATGG TGGGGGGGGC CCACTGGGGA GTCCTRGGG GCTTGGCTA CTATTCCATG 120
 GTRGGGAAGT GGGCTAAGGT TTGATTTGIG ATGCTACTCT TTGCGGGGT TGAOGGGMRT 180
 ACCCGGTGA OGGGGGGGT GCAAGGOCAY GTCACCTCTR CACTCAAGTC CCTCTTTAGA 240
 CCTGGGGGT CCCAGAAAAT TCAGYYTKTA AACACCAATG GCAGTTGGCA TATCAACAGG 300
 ACTGCGCTGA AYTGAATGA CTCCTCCAA ACTGGGTTCC TTGCGGCGCT G 351

40

(2) INFORMATION FOR SEQ ID NO:46:

45

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 583 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

55

EP 0 939 128 A2

CTCAGTGATC GACTGTAAAC CATGTGTAC TCAGAAGGTC GATTTGAGCT TGGATCCRC 60
 5 CTTCACCATC GAGAOGAOGA OOGTGCCCCA AGATGOGGTT TGGGCAOGC AGOGGOGAGG 120
 TAGGACTGGC AGGGGCAGGA GAGGCATCTA TAGGTTTGTG ACTCCAGGAG AAGGGCCCTC 180
 GGGGATGTTT GATTCTTGGG TCCTATGTGA GTGTTATGAC GGGGGMGTG CTGGGTATGA 240
 10 GCTCAGCCCC GCTGAGAOCT OGGTTAGGTT GGGGGCTTAC CTAAATACAC CAGGGTTGCC 300
 OGTCTGCCAG GACCATCTGG AGTCTGGGA GAGGCTCTC ACAGGCCTCA CCCACATAGA 360
 CGOCCACTTC TGTGCCAGA CTAAGCAGGC AGGAGACAAC TTCCCTACC TGGTAGCATA 420
 15 CCAAGOCACA GTGTGOGCA GGGCTAAGGC YOCACCTCA TGTGGGATC AAATGTGGAA 480
 GTGTCTCATA OGGCTAAAGC CTACGCTGCA OGGGCSAAG CCCTGCTGT ATAGRCTAGG 540
 20 AGOOGTCCAG AATGAGGTCA CCTCACACA CCTATAACC AAA 583

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
 25 (A) LENGTH: 427 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

OCTACCCGT GACCCACOG TCCCCCTGCG GGGGCTGGG TGGGAGACAG CTAGACACAC 60
 YCCAGTCAAC TCCTGGCTAG GCAACATCAT YATGTATGGG CCACCTTTGT GGGCAAGGAT 120
 35 GATTCTGATG ACTCACTTCT TCTCATCTCT TCTAGCCAG GAGCAACTTG AAAAAGCOCT 180
 GGATTGTCAA ATCTAOGGGG CCTGTTACTC CATTGAGCCA CTGACCTAC CTCAGATCAT 240
 40 TGAAGACTC CATGGTCTTA GGCATTTTC ACTCCATAGT TACTCTCCAG GTGAGATCAA 300
 TAGGGTGGCT TCATGCCTCA GGAAGCTTGG GGTACCAACC TTGGAGTCT GGAGACATGG 360
 GGOCAGAAGT GTCOGGCTA AGCTACTGTC CCARGGGGG AGGGCOGCCA CTTGTGGCAA 420
 45 GTACCTC 427

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
 50 (A) LENGTH: 552 base pairs
 (B) TYPE: nucleic acid

55

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

	AGCOGAGTAG TGTTGGGTG OGAAAGGCOCT TGIGGTACTG OCTGATAGGG TGCTTGOAG	60
10	TGCCCCGGGA GGTCTGCTAG ACOGTGCAYC ATGAGCACRA ATCTTAAACC YCAAAAAA	120
	AMCAAACGTA ACACCAACOG TGGCCACAG GACGTAAAGT TCCGGGYGG YGGTCAGATC	180
	GTGTTGGAG TTACTTGTG GCGGCGAGG GGCCYAGRT TGGGTGTGCG YGCGACKAGR	240
15	AAGACTTCOG AGGGGTGCA ACCTGCGGW AGCGWCARC CTATCCCAA GGCTGCGCG	300
	CCGAGGGCA GGACTGGGC TCAGCCYGGG TAYCCTGGC CCTCTATGG CAATGAGGGC	360
20	TKSGGTGGG CRGGATGGCT CCTGTWCC CGYGGCTCTC GGCTAGYTG GGGCCCCAMW	420
	GACCCCGGC GTAGGTGCG YAATTTGGGT AAGGTCATG ATACCTTAC RTGGGCTTC	480
	GCGACCTCA TGGGTACAT WCGCTYGTG GCGCCCCY TWGGRGGCG TGCCAGGGC	540
25	CTGGCRCATG GY	552

(2) INFORMATION FOR SEQ ID NO:49:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

	WCGCTGTC GCGCCCCY TWGGRGGCG TGCCAGGGC CTGGCRCATG GYGTCCGGT	60
40	TCTGGARGAC GGGTGAAC ATGCAACAGG GAAYTKCCY GGTGCTCTT TCTCTATCTT	120
	CCTYYTGGCY CTGCTSTCYT GYTTGACYRT SCMGCTTCS GCYTAYSAAG TGCGCAACKY	180
	SWCSGGMTW TACCAYGTCA CMAAYGAYTG CYCYAACTCR AGCATTGTGT AYGAGGGGGC	240
45	SGAYGYSATC MTGCAIRCYC CSGGGTGGT SCCYTGGTT CGKGAGRCA AYKCTCSMG	300
	KTYTGGGTG GGMTSACYC CYACGSTSGC SRCCAGGRAT GSCARMSTCC CCRCKACGMM	360
50	RYTWOGAGY CACRTGAYY TGCTYGTGG GASSGCYRCY YTCTGYTCG CYMTSTAOGT	420
	GGGGGAYCTM TGGGRTCTG TYTTYCTYRT CKSCCARCTG TTCACCTTCT CKCCYMGSG	480

55

EP 0 939 128 A2

SCAYKRGACR RYRCARGRYT GCAAYTGCTC WATCTATCCC GGOCAYRTAW CRGGYCAYCG 540
 CATGGCWTGG GATATGATGA TGAAC TGGTC SOCYACGRCR 580

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AACTGGTCSC CYACGRCRGC STTRGTTRRTG KCKCAGYTRC TCOGGATCCC ACAAGCYRTC 60
 WTGGACATGR TSGCKGGKGC YCACTGGGGA GTCTRGCGG GCMTWGCSTA YWYTTCCATG 120
 GTGGGGA ACT GGGCKAAGGT YYTGARTWTG MITGCTRCTMT TTGCCGGCGT YGAOGSGSAW 180
 ACCOCROGISA CSGGGGGRRK KSMGGOCAC RYRYSTCTR SAYTYRYKWS CCTCYTYRSA 240
 CCWGGSGCSW MSCAGAAMRT YCAGCTKRIM AACACCAAYG GCAGTTGGCA YMTCAAYAGS 300
 ACKGCOCTGA ACTGCAATGA YWSOCTOMAN ACYGGSTKSY TKGCMGSGCT K 351

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 583 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CTCRGTGATM GACTGYAAYA CRTGTGTAC YCAGACRGTC GATTTTCAGCY TKGAYCCYAC 60
 CTTCAACATY GAGACRAYSA CSSTSCCCCA RGATGCKGTY TCSGCKACKC AROGKOGRG 120
 YAGGACTGGC AGGGGSAREM SAGGCATCTA YAGRTTTGTG ROWCCRGGRG AROGSCCCTC 180
 SGSSATGTTT GAYTCKTCSG TCTMTGTGA GTGYTATGAC GCRGGCTGTG CTTGGTATGA 240
 GCTCAGCCCC GCYGAGACYW CRGTTAGGYT ROGRGCKTAC MIRAAYACMC CRGGGYTKCC 300

EP 0 939 128 A2

OGTSTGCCAG GAOCATCTKG ARTTYTGGGA GRGCGTCTTY ACAGGCCTCA CYCAYATAGA 360
 5 YGCCCCACTTY YTRTCCAGA CWAAGCAGRS WGGRGASAAC YTYCCYTACC TGGTAGCRTA 420
 CCAAGCCACM GTGTGCGCYA GGGCTMARGC YCCWCCYCCA TCGTGGGAYC ARATGTGGAA 480
 GIGTYTSATW OGSCIMAAGC CYACSTSCA YGGGCCAACR CCCCTGCIRT AYAGRCTRGG 540
 10 MGCYGTYCAG AATGARRTCA CCTSACRCA CCCWRIMACC AAA 583

15 (2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 427 base pairs
 20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

25 OCTCACCOTG GACCCYACMR YCCCCCTYGC GMERGCTGCG TGGGAGACAG CWAGACACAC 60
 TOCAGTCAAY TCCTGGCTAG GCAACATMAT CATGTWTCG COCACWYTGT GGGCRAGGAT 120
 30 GATWCTGATG ACYCAYTTCT TYTCCTCTCT TMTAGCCMRG GASCARCTTG AAMARGCCCT 180
 SGATTGYSAR ATCTAOGGGG CCTGYTACTC CATWGARCCA CTTGAYCTAC CTCMRATCAT 240
 TSAAMGACTC CATGGYCTYA GGCATTTTC ACTCCAYAGT TACTCTOCAG GTGARATYAA 300
 35 TAGGGTGGCY KCATGCCTCA GRAARCTTGG GGTACCRCCC TTGCGAGYTT GGAGACAYCG 360
 GGCCMRAGY GTCGOGCTA RGCTWCTGKC CMRAGGRGGS AGGGCYGCCA YWTGTGGCAA 420
 40 GTACCTC 427

45 (2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8865 base pairs
 (B) TYPE: nucleic acid
 50 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

55

EP 0 939 128 A2

	ATGAGCAOGA ATCCTAAACC TCAAAAAAAA AACAAAOGTA ACACCAACOG TOGCCCCACAG	60
5	GACGTCAAGT TCOOGGGTGG OGGTCAGATC GTTGGTGGAG TTTACTTGT TCOGOGCAGG	120
	GGCCCTAGAT TGGGTGTGOG OGGACGAGA AAGACTTCOG AGGGGTGCA AOCTOGAGGT	180
	AGAGTCAAGC CTATCCCCAA GGCTOGTOGG CCOGAGGGCA GGAOCTGGGC TCAGCCCCGG	240
10	TACCCCTGGC CCGTCTATGG CAATGAGGGC TGCGGGTGGG OGGGATGGCT CCTGTCTCCC	300
	CGTGGCTCTC GGCCTAGCTG GGGCCCCACA GACCCOOGGC GTAGGTGOG CAATTTGGGT	360
15	AAGGTCATOG ATACCCCTAC GTGCGGCTTC GCOGACCTCA TGGGGTACAT ACOGCTOGTC	420
	GGGCCCCCTC TTGGAGGOGC TGCCAGGGCC CTGGGCGATG GGTCCGGGT TCTGGAAGAC	480
	GGGGTGAAGT ATGCAACAGG GAACCTTCCT GGTGCTCTT TCTCTATCTT CCTTCTGGCC	540
20	CTGCTCTCTT GCTTGAAGT GCGCGCTTC GCTACCAAG TGCGCAACTC CAGGGGGCTT	600
	TACCAAGTCA CCAATGATTG CCCTAACTOG AGTATTGTGT ACGAGGGGCG CGATGOCATC	660
25	CTGCACACTC OGGGGTGGT CCGTTGGGTT CGTGAGGGCA ACGCCTOGAG GTGTTGGGTG	720
	GCGATGACCC CTACGGTGGC CACCAGGGAT GGCAAACTCC COGOGAOGCA GCTTCGACGT	780
	CACATOGATC TGCTTGTGCG GAGCGCCACC CTCTGTTGCG CCCTCTACGT GGGGGACCTA	840
30	TGCGGGTCTG TCTTTCTGTG OGGCCAACTG TTCACCTTCT CTCCCAGGCG CCCTGGAAG	900
	ACGCAAGGTT GCAATTGCTC TATCTATCCC GGCCATATAA OGGGTCAOOG CATGGCATGG	960
35	GATATGATGA TGAAGTGGT CCCTACGACG GCGTTGGTAA TGGCTCAGCT GCTCOGGATC	1020
	CCACAAGCCA TCTTGACAT GATCGCTGGT GCTCACTGGG GAGTCTGGC GGGCATAGOG	1080
	TATTTCTCCA TGGTGGGGAA CTGGGOGAAG GTCTGGTAG TGCTGCTGCT ATTTGCOGGC	1140
40	GTCGACGOGG AAACCCACGT CACCGGGGA AGTGCGGGC AACTGTGTG TGGATTGTGT	1200
	AGCCTCTCG CACCAGGCG CAAGCAGAAC GTCCAGCTGA TCAACACCA OGGCAGTTGG	1260
45	CACCTCAATA GCAOOGGCT GAACTGCAAT GATAGCCTCA ACACOGGCTG GTTGGCAGGG	1320
	TTTTCTATCA CCACAAGTTC AACTCTTCAG GCTGTCTGA GAGGCTAGOC AGCTGCOGAC	1380
	CCCTTACGA TTTTGACCAG GGCTGGGGCC CTATCAGTTA TGCCAACGGA AGOGGCCCCG	1440
50	ACCAGCGGCC CTACTGCTGG CACTACCCCC CAAAACCTTG OGGTATTGTG CCOGOGAAGA	1500
	GTGTGTGTGG TCGGTATAT TGCTTCACTC CCAGCCCCGT GGTGGTGGGA ACGACOGACA	1560
55		

EP 0 939 128 A2

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55

GGTOGGGOGC GCOACCTAC AGCTGGGGTG AAAATGATAC GGAOGTCTTC GTCCTTAACA 1620
 ATACCAGGCC AOCGCTGGGC AATTGGTTTG GTGTACCTG GATGAACTCA ACTGGATTCA 1680
 CCAAAGTGTG OGGAGOGGCT CCTTGIGTCA TOGGAGGGGC GGGCAACAAC AOCCTGCACT 1740
 GCOOACTIGA TTGCTTCCGC AAGCATOOGG AOGCCACATA CTCTGGGTGC GGCTOOGGTC 1800
 OCTGGATCAC AOCAGGTGC CTGGTOGACT AOCGTATAG GCTTTGGCAT TATCCTTGTA 1860
 CCATCAACTA CACCATATTT AAAATCAGGA TGTACGTGGG AGGGGTGAA CACAGGCTGG 1920
 AAGCTGCGTG CAACTGGAOG OGGGGOGAAC GTTGOATCT GGAAGACAGG GACAGGTCCG 1980
 AGCTCAGCCG GTTACTGCTG ACCACTACAC AGTGGCAGGT OCTCCOGTGT TCCCTCACAA 2040
 CCTTACCAGC CTGTCTCACC GGCTCATCC ACCTCCACCA GAACATTGTG GACGTGCAGT 2100
 ACTTGTACGG GGTGGGGTCA AGCATOOGT OCTGGGCCAT TAAGTGGGAG TACGTGTTTC 2160
 TCCGTCTCCT TCTGCTTGCA GAGGOGGCG TCTGCTCCTG CTTGTGGATG ATGCTACTCA 2220
 TATCCAAGC GGAGGOGGCT TTGGAGAACC TCGTAATACT TAATGCAGCA TCCCTGGCOG 2280
 GGAOGCAGG TCTGTATCC TTCCTGTGT TCTCTGCTT TGCATGGTAT TTGAAGGGTA 2340
 AGTGGGTGCC OGGAGOGGTC TACACCTTCT AOCGGATGTG GCTCTCTCTC CTGCTCTGT 2400
 TGGGTTTGGC CCAGOGGGOG TACGOGCTGG ACAGGAGGT GGCOGCTOG TGTGGGGTG 2460
 TTGTCTCTGT OGGGTGTATG GCGCTGACTC GTTACCATA TTACAAGGC TATATCAGCT 2520
 GGTGCTTGTG GTGGCTTCAG TATTTTCTGA CCAGAGTGA AGOGCAACTG CAOGTGTGA 2580
 TTCCCCCTT CAACTGCGA GGGGGGCGG AOCOGTCTAT CTTACTCATG TGTGCTGTAC 2640
 ACOGACTCT GGTATTTGAC ATCAACAAAT TGCTGCTGGC GGTCTTGGGA CCGCTTTGGA 2700
 TTCTTCAAGC CAGTTTGTCT AAAGTACCT ACCTTGTGG OGTCCAAGGC CTTCTOOGT 2760
 TCTGCGGTT AGOGCGGAAG ATGATGGAG GCCATTACGT GCAAATGGTC ATCATTAAGT 2820
 TAGGGGCGCT TACTGGCACC TATGTTTATA ACCATCTCAC TCCCTCTTGG GACTGGGGGC 2880
 ACAAGGCTT GOGAGATCTG GCGTGGCTG TAGAGCCAGT GGTCTTCTCC CAAATGGAGA 2940
 CCAAGCTCAT CAOGTGGGG GCAGATACG CCGGTGCGG TGACATCATC AACGGCTTGC 3000
 CTGTTTCCGC OGCAGGGGC OGGGAGATAC TGCTGGGCC AGCOGATGA ATGGTCTCCA 3060
 AGGGGTGGAG GTTGTGGOG CCGATCAGG GGTACGCCA GCAGACAAGG GGCTCTCTAG 3120

EP 0 939 128 A2

GGTGCATAAT CACCAGCCTA ACTGGGCGGG ACAAAAACCA AGTGGAGGGT GAGGTCCAGA 3180
5 TTGTGTCAAC TGCTGOCOA ACCTTCCTGG CAACGTGCAT CAATGGGGTG TGCTGGACTG 3240
TCTACCACGG GGCOCGAAOG AGGACCATOG OGTCACCCAA GGGTCCTGTC ATCCAGATGT 3300
ATACCAATGT AGACCAAGAC CTGTGGGGCT GGCOCGCTCC GCAAGGTAGC CGCTCATTGA 3360
10 CACCTGTCAC TTGCGGCTCC TOGGAOCTTT ACCTGGTCAC GAGGCAOGCC GATGTCATTG 3420
COGTGOGCOG GGGGGGTGAT AGCAGGGGCA GCGTCTGTC GOCOCGGGCC ATTTCTACT 3480
15 TGAAAGGCTC CTGGGGGGT COGCTGTGT GOCOCGGGG GCAOGCOGTG GGCATATTTA 3540
GGGCOGCGGT GTGCACCOGT GGAGTGGCTA AGGCGGTGGA CTTTATCCCT GTGGAGAAOC 3600
TAGAGACAAC CATGAGGTCC COGCTGTTC OGGATAACTC CTCCTCACCA GTAGTGOCCC 3660
20 AGAGCTTOCA GGTGGCTCAC CTCATGCTC CCACAGGCAG CGGCAAAAGC ACCAAGGTCC 3720
CGGCTGCATA TGCAGCTCAG GGCTATAAGG TGCTAGTACT CAACCCCTCT GTTGCTGCAA 3780
CACTGGGCTT TGGTGCTTAC ATGTCCAAGG CTCATGGGAT CGATCCTAAC ATCAGGACCG 3840
25 GGGTGAGAAC AATTACCACT GGCAGCCCCA TCAOGTACTC CACCTAOGGC AAGTTCTTGT 3900
COGAOGGOGG GTGCTOGGGG GGCGCTTATG ACATAATAAT TTGTGACGAG TGCCACTCCA 3960
30 CGGATGOCAC ATCCATCTTG GGCATOGGCA CTGTCTTGA CCAAGCAGAG ACTGOGGGGG 4020
CGAGACTGGT TGTGCTOGCC ACCGCCACCC CTCOGGGCTC CGTCACTGTG CCCCATCCA 4080
ACATOGAGGA GGTGTCTCTG TCCACCACCG GAGAGATCC TTTTAAOGGC AAGGCTATCC 4140
35 CCGTGAAGT AATCAAGGGG GGGAGACATC TCATCTTCTG TCATTCAAAG AAGAAGTGG 4200
ACGAACTOGC OGCAAAGCTG GTGCGATTGG GCATCAATGC CGTGGCCTAC TACOGGGTC 4260
40 TTGAOGTGTG CGTCATCCG ACCAGOGGCG ATGTTGTGT CGTGGCAACC GATGCCCTCA 4320
TGACCGGCTA TACOGGOGAC TTGACTOGG TGATAGACTG CAATACGTGT GTCACCCAGA 4380
CAGTOGATTT CAGCCTTGAC CCTACCTTCA CCATTGAGAC AATCAOGCTC CCCCAGGATG 4440
45 CTGTCTCCCG CACTCAAGT CGGGGCAGGA CTGGCAGGGG GAAGCCAGGC ATCTACAGAT 4500
TTGTGGCAAC GGGGGAGGCG CCGTCOGGCA TGTTCGACTC GTCGTCCTC TGTGAGTGCT 4560
50 ATGAOGCAGG CTGTGCTTGG TATGAGCTCA CGCCGCGCA GACTACAGTT AGGCTACGAG 4620
CGTACATGAA CACCCCGGGG CTTCCOGTGT GCCAGGACCA TCTTGAATTT TGGGAGGGCG 4680

55

EP 0 939 128 A2

5 TCCTTACAGG CCTCACTCAT ATAGATGCCC ACTTTCTATC CCAGACAAAG CAGAGTGGGG 4740
 AGAAOCTTCC TTACCTGGTA GGTACCAAG CCACCGTGTG CGCTAGGGCT CAAGCCOCTC 4800
 CCCCATGTG GGACAGATG TGAAGTGTG TGATTGCGCT CAAGCCCAAC CTCCATGGGC 4860
 CAACACCCCT GCTATACAGA CTGGGCGCTG TTCAGAATGA AATCACCTG ACGCAOCCAG 4920
 10 TCACCAAATA CATCATGACA TGCATGTGG CGACCTGGA GGTCGTCAAG AGCACCTGGG 4980
 TGCTGTGTGG OGGGTGCTG GCTGCTTTGG CCGGTATTG CCTGTCAACA GGCTGTGTGG 5040
 TCATAGTGGG CAGGTGTGTC TTGTCCGGGA AGCCGGCAAT CATACCTGAC AGGGAAGTCC 5100
 15 TCTACCGAGA GTTCGATGAG ATGGAAGAGT GCTCTCAGCA CTTACCGTAC ATCGAGCAAG 5160
 GGATGATGCT CGCGAGCAG TTCAAGCAGA AGGCOCTGG CCTCCTGCAG ACCGGTCCC 5220
 20 GTCAGGCAGA GGTATGGCC CCGTGTGTCC AGACCAACTG GCAAAACTC GAGACCTTCT 5280
 GGGGGAAGCA TATGTGGAAC TTCATCAGTG GGATACAATA CTGGGGGGC TTGTCAACGC 5340
 TGCTGTGTA CCGGOCATT GCTTCATTGA TGGCTTTTAC AGCTGTGTG ACCAGCCAC 5400
 25 TAAOCTAG CCAACCCCTC CTCTTCAACA TATTGGGGG GTGGGTGGCT GCCAGCTOG 5460
 CCGCCCCGG TCGGCTACT GCTTTGTGG GCGTGGCTT AGCTGGGCG GGCATGGCA 5520
 30 GTGTGGACT GGGGAAGTC CTCATAGACA TCGTGCAGG GTATGGGCG GGGTGGCGG 5580
 GAGCTCTGT GGCATTCAAG ATCATGAGG GTGAGGTCCC CTCCACGAG GACCTGGTCA 5640
 ATCTACTGCG CGCCATCCTC TCGCCCGAG CCTGTAGT OGGGTGTGTC TGTGCAGCA 5700
 35 TACTGCGCG GCAOCTTGG CCGGGGAGG GGCAGTGCA GTGGATGAAC CGCTGATAG 5760
 CCTGTGCTC CCGGGGAAC CATGTTTCCC CACGCACTA CGTGGCGAG AGCGATGCAG 5820
 40 CTGCGCGGT CACTGCCATA CTCAGCAGC TCACTGTAAC CCAGCTCTG AGGCGACTGC 5880
 ACCAGTGGAT AAGCTCGAG TGTACCACTC CATGCTCGG TTCTGGCTA AGGACATCT 5940
 GGGACTGGAT ATCGAGGTG TTGAGCGACT TTAAGACCTG GCTAAAAGCT AAGCTCATGC 6000
 45 CACAGCTGCG TGGGATCCC TTGTGTCTC GCCAGCGCG GTATAAGGGG GTCTGGCGAG 6060
 TGGACGGCA TCATGCACAC TCGCTGCCA TGTGGAGCTG AGATCACTGG ACATGTCAAAA 6120
 50 ACGGACGAT GAGGATGTC GTCTTAGGA CCGCAGGAA CATGTGGAGT GGGACCTCC 6180
 CCATTAATGC CTACACCAAG GGGCCCTGTA CCCCCCTCC TGGCGGAAC TACAGTTGG 6240

55

EP 0 939 128 A2

	CGCTATGGAG GGTGTCTGCA GAGGAATATG TGGAGATAAG GCAGGTGGGG GACTTCCACT	6300
5	AOGTGAOGGG TATGACTACT GACAATCTCA AATGCCOOGT CCAGGTCCCA TOGCCOGAAT	6360
	TTTTACAGA ATTGGACGGG GTGGOGCTAC ATAGGTTTGC GCOOCCCTGC AAGCCCTTGC	6420
	TGOGGGAGGA GGTATCATTG AGAGTAGGAC TOCAOGAATA CCOGGTAGGG TOGCAATTAC	6480
10	CTTGGAGAGC OGAACCGGAC GTGGCOGTGT TGACGTCCAT GCTCACTGAT CCTCCCATTA	6540
	TAACAGCAGA GGOGGCOGGG CGAAGGTITG OGAGGGGATC ACCOCCCTCT GTGGCCAGCT	6600
	CCTGGGCTAG CCAGCTATCC GCTCCATCTC TCAAGGCAAC TTGCACCGCT AACCATGACT	6660
15	CCCTGATGC TGAGCTCATA GAGGCOCAACC TCCTATGGAG GCAGGAGATG GGCGGCAACA	6720
	TCACCAGGGT TGAGTCAGAA AACAAAGTGG TGATTCTGGA CTCTTTOGAT CCGCTTGTGG	6780
20	CGGAGGAGGA CGAGOGGGAG ATCTCOGTAC CGCAGAAAT CCTGOGGAAG TCTOGGAGAT	6840
	TOGCCAGGC CCTGCCOGTG TGGGOGCGGC CGGACTATAA CCCCCCGCTA GTGGAGACGT	6900
	GGAAAAAGCC CGACTACGAA CCACCTGTGG TCATGGCTG TCCGCTTCCA CCTCCAAAGT	6960
25	CCCTCTCTGT GCTCCOCTCT CGGAAGAAGC GGAOGGTGGT CCTCACTGAA TCAACCTAT	7020
	CTACTGCTTT GGOOGAGCTC GOCACCAGAA GCTTTGGCAG CTCTCAACT TOGGCATTA	7080
30	CGGGGACAA TAAGACAACA TCCTCTGAGC CGCCCCCTTC TGGCTGCCCC CCGACTCCG	7140
	AOGCTGAGTC CTATTCTCTC ATGCCOCCCC TGGAGGGGGA GCCTGGGGAT CCGGATCTTA	7200
	GCGAOGGGTC ATGGTCAACG GTCAGTAGTG AGGCCAACGC GGAGGATGTC GTGTGCTGCT	7260
35	CAATGTCTTA CTCTTGGACA GGCGCACTOG TCACCCCGTG CGCGGGGAA GAACAGAAAC	7320
	TGCCCATCAA TGCACTAAGC AACTOGTTGC TAGCTACCA CAATTTGGTG TATTOACCA	7380
40	CCTCAOGCAG TGCTTGCCAA AGGCAGAAGA AAGTCACATT TGACAGACTG CAAGTTCTGG	7440
	ACAGCCATTA CCAGGAAGTA CTCAAGGAGG TTAAAGCAGC GGCGTCAAAA GTGAAGGCTA	7500
	ACTTGCTATC CGTAGAGGAA GCTTGAGGCC TGACGCCCC ACACTCAGCC AAATCCAGT	7560
45	TTGGTTATGG GGCAAAAGAC GTCCGTTGCC ATGCCAGAAA GGCGTAACC CACATCAACT	7620
	CCGTGTGGAA AGACCTTCTG GAAGACAATG TAACACCAAT AGACACTAOC ATCATGGCTA	7680
50	AGAOCGAGGT TTCTTGCTT CAGCCTGAGA AGGGGGGTGG TAAGCCAGCT CGTCTCATCG	7740
	TGTTCCCGA TCTGGGCGTG CGGTGTGCG AAAAGATGCC TTTGTACGAC GTGGTTACAA	7800

55

EP 0 939 128 A2

AGCTCCCCCTT GGOGTGATG GGAAGCTCCT ACGGATTCCA ATACTCACCA GGACAGGGG 7860
5 TTGAATTCTT CGTGCAAGCG TGGAAGTCCA AGAAAACCCG AATGGGGTTC TCGTATGATA 7920
OOOGCTGCTT TGACTCCACA GTCACTGAGA GOGACATCG TAOGGAGGAG GCAATCTACC 7980
AATGTTTGTA COTGACCCG CAAGCCCGCG TGGCCATCAA GTCCCTCACC GAGAGGCTTT 8040
10 ATGTTGGGGG CCTCTTACC AATTCAAGGG GGGAGAACTG CGGCTATCGC AGGTGCGCGG 8100
CGAGCGGGGT ACTGACAACT AGCTGTGGTA ACACCCCTCAC TTGCTACATC AAGGCCCGGG 8160
CAGCCTGTG AGCGCGAGGG CTCCAGGACT GCACCATGCT CGTGTGTGGC GAGGACTTAG 8220
15 TCGTATCTG TGAAAGCGCG GGGGTCCAGG AGGACGCGGC GAGCCTGAGA GCGTTCAAGG 8280
AGGCTATGAC CAGGTACTOC GCGCCCGCTG GGGACCCCGC ACAACCAGAA TAGACTTGG 8340
AGCTCATAAC ATCATGCTCC TCCAACTGT CAGTCGCCA CGACGGCGCT GGAAAGAGGG 8400
TCTACTACCT CACCGTGAC CTTACAACC CCTCGCGAG AGCTGCGTGG GAGACAGCAA 8460
GACACACTOC AGTCAATTCC TGGCTAGGCA ACATAATCAT GTTTGCCCC ACCTGTGGG 8520
25 CGAGGATGAT ACTGATGACC CATTTCTTTA GGTCTCTAT AGCCAGGGAC CAGCTTGAAC 8580
AGGCGCTGTA TTGCGAGATC TAGGGGCGT GCTACTCCAT AGAACCCTT GATCTACCTC 8640
CAATCATTCA AAGACTOCAT GGCTCAGCG CATTTTCACT CCACAGTTAC TCTCCAGGTG 8700
30 AAATTAATAG GGTGGCGCA TGCTCAGAA AACTTGGGGT ACGCCCTTG CGAGCTTGA 8760
GACACCGGGC CCGGAGGTC CGGCTAGGC TTCTGCCAG AGGAGGCAGG GCTGCCATAT 8820
35 GTGGCAAGTA CCTCTTCAAC TGGGCAGTAA GAACAAAGCT CAAAC 8865

(2) INFORMATION FOR SEQ ID NO:54:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 367 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

YWSOCTCAAM ACYGGSTKKY TKGCGSGCT KTTCTAYMM CACAAGTTCA ACKCKTOMGG 60
50 MITGYCKGAG MGSMTGCCA GCTGYGMYC CMITRMCAAK TTYGACCAGG GMTGGGGYCC 120

55

YATCASYTAT GCYMAMSSWR RCRRCYCSGA CCAGMGSOCs TAYTGCTGGC ACTACSCMOC 180
 5 WMRACMKIGY GGTATYGTIC COGOGMMEMR KGTGTGYGGT CCRGTIRTATT GCTTCACYOC 240
 MAGCCCYGK GIRGTGGRA OGACOGAYMG KTYSGGOGCS CCYACSTAYA RCTGGGGKGA 300
 MAATGAKAG GAOGTSYTSS TCCTWAACAA YACSMGGCOM COGWSGGCA AYTGGTTOGG 360
 10 YTGTA 367

(2) INFORMATION FOR SEQ ID NO:55:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1249 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

WCGCTOGTC GGGGCCCCYY TWGGRGGGOC TGCCAGGGOC CTGGCRCATG GYGTCCGGGT 60
 25 TCTGGARGAC GGGGTGAAC ATGCAACAGG GAAYTKCCY GGTGTCTCTT TCTCTATCTT 120
 CCTYYTGGCT CTGCTSTCYT GYTTGACYRT SCCMGCTTCS GCYTAYSAG TGCGCAACKY 180
 SWCSGGGMIW TACCAYGTCA CMAAYGAYTG CYCYAACTCR AGYATTGTGT AYGAGGGGGC 240
 30 SGAYGYSATC MITCAVRCYC CSGGGTGOGT SCCYTGGTT CGKGAGRCA AYKCTCSMG 300
 KTGTYGGGTA GCGMTSACYC CYAGSTSGC SRCCAGGRAT GSCARMSTOC CCRCIAOGMM 360
 35 RYTWGAGCY CACRTGAVY TGCTYGTYGG GASSGCYRCY YTCTGYTCSG CYMTSTAOGT 420
 GGGGAYCIM TGGGRTCTG TYTTYCTYRT CKSCCARCTG TTCACCTTCT CKCCYMGSG 480
 SCAYKRGACR RYRCARGRYT GCAAYTGCTC WATCTATCCC GGCCAYRTAW CRGGYCAVCG 540
 40 CATGGCWTTG GATATGATGA TGAAGTGGT SCCYAAGRRC GCSTTRGIRR TGKCKCAGYT 600
 RCTOOGGATC CCACAAGCYR TCWTGGACAT GRTSGCKGGK GCYCACTGGG GAGTCCTRGC 660
 GGGGMIWGS TAYTWYTOCA TGGTGGGGAA CTGGGCKAAG GTYYTGRTWG TGMTGCTRCT 720
 45 MITTGCCGGC GTYGAOGSGS AWACCCROGT SACSGGGGGR RKKSMGGCC ACRYRYSTC 780
 TRSAYTYRYK WSCCTCYTYR SACCWGGSGC SWMSCAGAAM RYTCAGCTKR TMAACACCAA 840
 50 YGGCAGTTGG CAYMTCAAYA GSACKGCCCT GAACTGCAAT GAYWSCCTOM AMACYGGSTK 900

EP 0 939 128 A2

SYTRGOMGSG CTKTCTAYM MMCACAAGTT CAACKCKTCM GGMIGYOCKG AGMGSMTGRG 960
 5 CAGCTGYOGM YCOMITRMCR AKTTYGACCA GGGMTGGGGY CCYATCASVT ATGCMYAMSS 1020
 WRRRCRCYCS GACCAGMGS CSTATYTGCTG GCACTACSCM COWMRAOMKT GYGGTATYGT 1080
 RCOOGOGWMG MRKGTGTGYG GTCCRGTRTA TIGCTTCACY COMAGCCCYG TKGTRGTGGG 1140
 10 RACGACOGAT MEKTYSGGOG CSCCYACSTA YARCTGGGGK GAMAATGAKA OGGAOGTSYT 1200
 SSTOCTWAAC AAYACSMGGC CMCOGWSGG CAAYTGGTTC GGYTGTACA 1249

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 278 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

25 TGGGCAAYTG GTTCGGYTGT ACMTGGAATGA AYWSMACTGG RTTCACCAAR RYGTGOGGAG 60
 SSCCYOCKTG TRWCATOGGR GGGYSGGCA ACAACACCYT GMMCTGCCCC ACKGAYTGCT 120
 TCOGSAAGMM YCCGRMSGOC ACYTACWWM RRTGYGGYT C SGGYCCYTGG WTSACACCYA 180
 30 GGTGCTYGGT YGACTACOCR TAYAGGCTYT GGCAYTAYCC YTGACYRTC AACTWYACCA 240
 TMTTYAARRT YAGGATGTAY GTGGGRGGSG TSGARCAC 278

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1539 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

45 ACACTGGGCT TTGGTGCTWA YATGTCCAAG GCWCATGGS A YOGAYCCYAA CATCAGRACY 60
 GGGGTRAGRA CMATYACCAC WGGYRSCCCC ATYACGTACT CCACCTAYSG CAAGTTCCTT 120
 GCOGACGGYG GKTGCTCSGG GGGGCGYTAT GACATMATAA TTTGTGACGA GTGCCACTCC 180
 50 ACGGATGCCA CATCCATCTT GGCATCGGC ACTGTCTCTG ACCAAGCAGA GACTGCGGGG 240

	GOGAGACTGG TTGTGCTOGC CACCGCCACC CCTCOGGGCT COGTCACTGT GCCCCATCCC	300
5	AACATOGAGG AGGTTGCTCT GTCCACCACC GGAGAGATCC CTTTTTAOGG CAARRSYATC	360
	CCCMTOGARG YMATCAAGGG GGRAGRCAT CTCATCTTCT GYCATTCTMAA GAAGAAGTGY	420
	GAGARCTOG COGCAAAGCT GKYMGCMTS GGMTCAATG COGTGGCSTA YTAOCGOGGT	480
10	CTTGAVGTGT COGTCAIMCC RACYAGOGGM GAYGTYGTYG TCGTGGCAAC MGAYGOOCTC	540
	ATGACOGGCT ATACOGGOGA CTTGACTOG GTGATAGACT GCAATACGTG TGTCACCCAG	600
15	ACAGTOGATT TCAGCCTTGA COCTACCTTC ACCATTGAGA CAATCAOGCT CCCCCAGGAT	660
	GCTGTCTOOC GCACTCAACG TOGGGGCAGG ACTGGCAGGG GGAAGCCAGG CATCTACAGA	720
	TTGTGGCAC CGGGGGAGOG COOCTCOGGC ATGTTGACT CGTCOGTCT CTGTGAGTGC	780
20	TATGAOCGAG GCTGTGCTTG GTATGAGCTC ACGCCOCCOG AGACTACAGT TAGGCTACGA	840
	GOGTACATGA ACACCCOGGG GCTTCCOGTG TGCCAGGACC ATCTTGAATT TTGGGAGGGC	900
25	GTCTTTACAG GOCTCACTCA TATAGATGCC CACTTTCTAT CCCAGACAAA GCAGAGTGGG	960
	GAGAACCTTC CTTAOCCTGT AGOGTACCAA GCCACOGTGT GOGCTAGGGC TCAAGCCCTT	1020
	CCCCATOGT GGGACCAGAT GTGGAAGTGT TTGATTGCC TCAAGCCAC CCTCCATGGG	1080
30	CCAACACCCC TGCTATACAG ACTGGGCGCT GTTCAGAATG AAATCACCTT GACGCACCCA	1140
	GTCAACAAAT ACATCATGAC ATGCAITGCG GCGACCTGG AGGTGCTCAC GAGCACCTGG	1200
35	GTGCTOGTTG GGGGCTOCT GGCTGCTTTG GCGOGTATT GCTGTCAAC AGGCTGOGTG	1260
	GTCTAGTGG GCAGGGTGT CTTGTCCGGG AAGCCGGCAA TCATACCTGA CAGGGAAGTC	1320
	CTCTAOCGAG AGTTOGATGA GATGGAAGAG TGCKCYVMRC ACYTMCCSTA CATOGARCAR	1380
40	GGRATGMWGC TOGCOGAGCA RTTCAAGCAG AAGGCSCTCG GSYTSCTGCA RACMGCSWCC	1440
	MRKCARGCRG AGGYTRYVGC YCCKKSTGWS YMRAYSMACK SSYMRAAACT CGAGACCTTC	1500
45	TGGGCGAAGC ATATGTGGAA CTTTCATCAGT GGGATACAA	1539

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 341 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

EP 0 939 128 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

5	GYTGGGCAAT TGGTTOGGTT GYACCTGGAT GAACTCAWCT GGATTYACCA AAGTGTGCGG	60
	AGOGCTOCT TGTGTCATCG GAGGGGYGGG CAACAACACC YTGCMITGCC CCACTGAYTG	120
	TTTCGGCAAG CATCOGGAAG CCACATACTC TCGGTGCGGY TCOGGTTOCT GGATYACRCC	180
10	CAGGTGCTG GTCSACTACC CKTATAGGCT TTGGCATTAT CCYTGTACYR TCAACTACAC	240
	CWERTTYAAA RUCAGGATGT ACGTGGGAGG GGTGARCAC AGGCTGGAAG YTGCTGCAA	300
15	CTGGAAGGGG GGGAROGTT GYGATCTGGA MGACAGGGAC A	341

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- | | |
|----|----------------------------|
| 20 | (A) LENGTH: 293 base pairs |
| | (B) TYPE: nucleic acid |
| | (C) STRANDEDNESS: double |
| | (D) TOPOLOGY: linear |

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

25	ATGAGCACRA ATCTAAACC TCAARAAAA AMCAAACGTA ACACCAACCG YGCCCCACAG	60
	GAAGTCAAGT TCOGGGYGG YGGTCAGATC GTTGGTGGAG TTTACTGTGT GCOGCGCAGG	120
30	GGCCCYAGRT TGGGTGTGCG CGGACKAGR AAGACTTCOG AGGGTTCGCA ACCTGWWGW	180
	AGGCGWCARC CTATCCCCAA GGCTGVCRG CCGAGGGCA GRCCTGGGC TCAGCCCGGG	240
35	TACCTTGGC CCTCTATGG CAAYGAGGGC WKSGGGTGG CRGGATGGCT OCT	293

40 Claims

1. A polynucleotide in substantially isolated form comprising a nucleotide sequence of at least 15 nucleotides from a J-7 HCV isolate, said J-7 HCV isolate having at least 90% nucleotide sequence homology with the J-7 sequence of Figure 1 or 6, wherein said nucleotide sequence of at least 15 nucleotides is distinct from the nucleotide sequence of HCV isolate HCV-1 as shown in Figure 12.
2. A polynucleotide according to claim 1 wherein the J-7 HCV isolate has at least 95% homology with the J-7 HCV sequence of Figure 1 or 6.
3. A polynucleotide according to claim 1 wherein the J-7 HCV isolate has 100% homology with the J-7 HCV sequence of Figure 1 or 6.
4. A polynucleotide according to any one of the preceding claims which comprises at least 20 nucleotides.
5. A method of detecting HCV polynucleotides in a test sample comprising:
 - (a) providing a polynucleotide as defined in any one of claims 1 to 4 as a probe;
 - (b) contacting the test sample and the probe under conditions that allow for the formation of a polynucleotide

EP 0 939 128 A2

duplex between the probe and its complement in the absence of substantial polynucleotide duplex formation between the probe and non-HCV polynucleotide sequences present in test sample; and

(c) detecting any polynucleotide duplexes comprising the probe.

- 5 6. A polynucleotide comprising a sequence of at least 15 nucleotides from a J-7 HCV isolate present in any one of plasmids pS1-8791a, bU1-1216c, bU1-4652d, pS1-713c, pS7-28c, pS1-1519, TC-600BP, JH-400BP, AW-300BP, AW-770-BP-N, AW-700BP-C/ AW-700BP-N or J1 5-1-1 deposited under accession numbers BP-2593, BP-2594, BP-2595, BP-2637, BP-2638, BP-3081, 68393, 68394, 68392, 68395 and 40884 respectively, wherein said nucleotide sequence of at least 15 nucleotides is distinct from the nucleotide sequence of HCV isolate HCV-1.
- 10 7. A purified polypeptide comprising a amino acid sequence which:
- (a) is encoded by a nucleotide sequence as defined in any one of claims 1 to 4 or in the HCV sequences deposited and defined in claim 6, said coding being in frame with the corresponding amino acid sequences set out in Figures 1 and 6.
- 15 (b) comprises an antigenic determinant; and
- (c) is distinct from the sequence of the polypeptides encoded by the HCV isolate HCV-1.
8. A polypeptide according to claim 7 which comprises at least 10 amino acids.
- 20 9. A polypeptide according to claim 7 which comprises at least 15 amino acids.
10. A polypeptide according to any one of claims 7 to 9 immobilised on a solid support.
- 25 11. An immunoassay for detecting the presence of anti-HCV antibodies in a test sample which comprises:
- (a) incubating the test sample under conditions that allow the formation of an antigen-antibody complex to be formed with a polypeptide as defined in any one of claims 7 to 10, wherein the polypeptide is not immunologically cross-reactive with HCV-1; and
- 30 (b) detecting any antigen antibody complexes formed.
12. An immunoassay according to claim 11 wherein the test sample comprises human blood or a fraction thereof.

J7 1 AGCCGAGTAGTGTGGGTCGCGAAAGGCCTTGTGGT
discrepancy
clone
altered aa

J7 37 ACTGCCTGATAGGGTGCTTGGCAGTGCCCCGGGAGG

J7 73 TCTCGTAGACCGTGCATC ATG AGC ACA AAT
Met Ser Thr Asn

J7 103 CCT AAA CCC CAA AGA AAA ACC AAA CGT
Pro Lys Pro Gln Arg Lys Thr Lys Arg
T G
b 1
--- Arg

J7 130 AAC ACC AAC CGT CGC CCA CAG GAC GTT
Asn Thr Asn Arg Arg Pro Gln Asp Val
C
b

J7 157 AAG TTC CCG GGC GGT GGT CAG ATC GTC
Lys Phe Pro Gly Gly Gly Gln Ile Val
T
1
Leu
T
b

J7 184 GGT GGA GTT TAC TTG TTG CCG CGC AGG
Gly Gly Val Tyr Leu Leu Pro Arg Arg
A
b

J7 211 GGC CCC AGG TTG GGT GTG CGT GCG ACT
Gly Pro Arg Leu Gly Val Arg Ala Thr

FIG. 1-1

J7	238	Arg Lys Thr Ser Glu Arg Ser Gln Pro
		AGG AAG ACT TCC GAG CGG TCG CAA CCT
		A
		b

J7	265	Arg Gly Arg Arg Gln Pro Ile Pro Lys
		CGT GGA AGG CGA CAA CCT ATC CCC AAG
J7	292	Ala Arg Arg Pro Glu Gly Arg Thr Trp
		GCT CGC CGG CCC GAG GGC AGG ACC TGG
J7	319	Ala Gln Pro Gly Tyr Pro Trp Pro Leu
		GCT CAG CCT GGG TAT CCT TGG CCC CTC
J7	346	Tyr Gly Asn Glu Gly Leu Gly Trp Ala
		TAT GGC AAT GAG GGC TTG GGG TGG GCA
		A
		b
		END
J7	373	Gly Trp Leu Leu Ser Pro Arg Gly Ser
		GGA TGG CTC CTG TCA CCC CGC GGC TCT
J7	400	Arg Pro Ser Trp Gly Pro Asn Asp Pro
		CGG CCT AGT TGG GGC CCC AAT GAC CCC
		T C
		c b
		--- Thr
J7	427	Arg Arg Arg Ser Arg Asn Leu Gly Lys
		CGG CGT AGG TCG CGT AAT TTG GGT AAG
J7	454	Val Ile Asp Thr Leu Thr Cys Gly Phe
		GTC ATC GAT ACC CTT ACA TGC GGC TTC
		C
		l
		Leu

FIG. 1-2

J7	481	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Leu
		GCC	GAC	CTC	ATG	GGG	TAC	ATT	CCG	CTT
								c		c
								c		b
								---		---
J7	508	Val	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala
		GTC	GGC	GCC	CCC	TTA	GGG	GGC	GCT	GCC
J7	535	Arg	Ala	Leu	Ala	His	Gly			
		AGG	GCC	CTG	GCA	CAT	GGT			

FIG. 1-3

J1 1 T Pro Leu Val Gly Ala Pro Leu Gly Gly
 discrepancy CCG CTC GTC GGC GCC CCC TTA GGG GGC
 clone C
 altered aa d
 Ser

29 Ala Ala Arg Ala Leu Ala His Gly Val
 GCT GCC AGG GCC CTG GCA CAT GGT GTC

56 Arg Val Leu Glu Asp Gly Val Asn Tyr
 CGG GTT CTG GAG GAC GGC GTG AAC TAT

83 Ala Thr Gly Asn Leu Pro Gly Cys Ser
 GCA ACA GGG AAT TTG CCC GGT TGC TCT

110 Phe Ser Ile Phe Leu Leu Ala Leu Leu
 TTC TCT ATC TTC CTC TTG GCT CTG CTG
 A T
 g d
 --- ---

137 Ser Cys Leu Thr Ile Pro Ala Ser Ala
 TCC TGT TTG ACC ATC CCA GCT TCC GCT

164 Tyr Glu Val Arg Asn Val Ser Gly Ile
 TAT GAA GTG CGC AAC GTG TCC GGG ATA

191 Tyr His Val Thr Asn Asp Cys Ser Asn
 TAC CAT GTC ACA AAC GAC TGC TCC AAC
 T
 d

218 Ser Ser Ile Val Tyr Glu Ala Ala Asp
 TCA AGC ATT GTG TAT GAG GCG GCG GAC

FIG. 2-1

245	Val Ile Met His Ala Pro Gly Cys Val
	GTG ATC ATG CAT GCC CCC GGG TGC GTG
272	Pro Cys Val Arg Glu Asn Asn Ser Ser
	CCC TGC GTT CGG GAG AAC AAT TCC TCC
	C
	d

299	Arg Cys Trp Val Ala Leu Thr Pro Thr
	CGT TGC TGG GTA GCG CTC ACT CCC ACG
326	Leu Ala Ala Arg Asn Ala Ser Val Pro
	CTC GCG GCC AGG AAT GCC AGC GTC CCC
353	Thr Thr Thr Leu Arg Arg His Val Asp
	ACT ACG ACA TTA CGA CGC CAC GTC GAC
	G
	d

380	Leu Leu Val Gly Thr Ala Ala Phe Cys
	TTG CTC GTT GGG ACG GCT GCT TTC TGC
407	Ser Ala Met Tyr Val Gly Asp Leu Cys
	TCC GCT ATG TAC GTG GGG GAT CTC TGC
434	Gly Ser Val Phe Leu Ile Ser Gln Leu
	GGA TCT GTT TTC CTC ATC TCC CAG CTG
	T
	d

461	Phe Thr Phe Ser Pro Arg Arg His Glu
	TTC ACC TTC TCG CCT CGC CGG CAT GAG

FIG. 2-2

488 Thr Val Gln Asp Cys Asn Cys Ser Ile
ACA GTA CAG GAC TGC AAC TGC TCA ATC

515 Tyr Pro Gly His Val Ser Gly His Arg
TAT CCC GGC CAC GTA TCA GGC CAT CGC
 T
 C

542 Met Ala Trp Asp Met Met Met Asn Trp
ATG GCT TGG GAT ATG ATG ATG AAC TGG

569 Ser Pro Thr Ala
TCG CCC ACG GCA

FIG. 2-3

J1		1		Asn Trp Ser Pro Thr AAC TGG TCG CCC ACG						
discrepancy										
clone										
altered aa										
J1	16	Ala GCA !!!	Ala GCC	Leu TTA	Val GTG	Val GTG	Ser TCG	Gln CAG	Leu TTA	Leu CTC
J1	43	Arg CGG	Ile ATC	Pro CCA	Gln CAA	Ala GCT	Val GTC	Met ATG	Asp GAC	Met ATG
J1	70	Val GTG	Ala GCG	Gly GGG	Ala GCC	His CAC	Trp TGG	Gly GGA	Val GTC	Leu CTA G i ---
J1	97	Ala GCG	Gly GGC	Leu CTT	Ala GCC	Tyr TAC	Tyr TAT	Ser TCC	Met ATG	Val GTG A i ---
J1	124	Gly GGG	Asn AAC	Trp TGG	Ala GCT	Lys AAG	Val GTT	Leu TTG	Ile ATT	Val GTG
J1	151	Met ATG	Leu CTA	Leu CTC	Phe TTT	Ala GCC	Gly GGC	Val GTT	Asp GAC	Gly GGG
J1	178	His CAT AG gg Ser	Thr ACC	Arg CGC	Val GTG	Thr ACG	Gly GGG A i ---	Gly GGG	Val GTG	Gln CAA

FIG. 3-1

J1	205	Gly	His	Val	Thr	Ser	Thr	Leu	Thr	Ser
		GGC	CAC	GTC	ACC	TCT	ACA	CTC	ACG	TCC
			T				G			
			C				i			
			---				Ala			
J1	232	Leu	Phe	Arg	Pro	Gly	Ala	Ser	Gln	Lys
		CTC	TTT	AGA	CCT	GGG	GCG	TCC	CAG	AAA
J1	259	Ile	Gln	Leu	Val	Asn	Thr	Asn	Gly	Ser
		ATT	CAG	CTT	GTA	AAC	ACC	AAT	GGC	AGT
				TC	T					
				ii	i					
				Ser	Leu					
J1	286	Trp	His	Ile	Asn	Arg	Thr	Ala	Leu	Asn
		TGG	CAT	ATC	AAC	AGG	ACT	GCC	CTG	AAC
										T
										g

J1	313	Cys	Asn	Asp	Ser	Leu	Gln	Thr	Gly	Phe
		TGC	AAT	GAC	TCC	CTC	CAA	ACT	GGG	TTC
J1	340	Leu	Ala	Ala	Leu					
		CTT	GCC	GCG	CTG					

FIG. 3-2

FIG. 4-1

J1	248	Pro	Ala	Glu	Thr	Ser	Val	Arg	Leu	Arg
		CCC	GCT	GAG	ACC	TCG	GTT	AGG	TTG	CGG
J1	275	Ala	Tyr	Leu	Asn	Thr	Pro	Gly	Leu	Pro
		GCT	TAC	CTA	AAT	ACA	CCA	GGG	TTG	CCC
J1	302	Val	Cys	Gln	Asp	His	Leu	Glu	Phe	Trp
		GTC	TGC	CAG	GAC	CAT	CTG	GAG	TTC	TGG
J1	329	Glu	Ser	Val	Phe	Thr	Gly	Leu	Thr	His
		GAG	AGC	GTC	TTC	ACA	GGC	CTC	ACC	CAC
J1	356	Ile	Asp	Ala	His	Phe	Leu	Ser	Gln	Thr
		ATA	GAC	GCC	CAC	TTC	TTG	TCC	CAG	ACT
J1	383	Lys	Gln	Ala	Gly	Asp	Asn	Phe	Pro	Tyr
		AAG	CAG	GCA	GGA	GAC	AAC	TTC	CCC	TAC
J1	410	Leu	Val	Ala	Tyr	Gln	Ala	Thr	Val	Cys
		CTG	GTA	GCA	TAC	CAA	GCC	ACA	GTG	TGC
J1	437	Ala	Arg	Ala	Lys	Ala	Pro	Pro	Pro	Ser
		GCC	AGG	GCT	AAG	GCT	CCA	CCT	CCA	TCG
						C				
						e				
						Ala(=)				
J1	464	Trp	Asp	Gln	Met	Trp	Lys	Cys	Leu	Ile
		TGG	GAT	CAA	ATG	TGG	AAG	TGT	CTC	ATA
J1	491	Arg	Leu	Lys	Pro	Thr	Leu	His	Gly	Pro
		CGG	CTA	AAG	CCT	ACG	CTG	CAC	GGG	CCA
									G	
									e	
									Ala	

FIG. 4-2

J1	518	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala
		ACG	CCC	CTG	CTG	TAT	AGG	CTA	GGA	GCC
							A			
							e			
							Arg(-)			
J1	545	Val	Gln	Asn	Glu	Val	Thr	Leu	Thr	His
		GTC	CAG	AAT	GAG	GTC	ACC	CTC	ACA	CAC
J1	572	Pro	Ile	Thr	Lys					
		CCT	ATA	ACC	AAA					

FIG. 4-3

J1 1
discrepancy
clone
altered aa

Leu Thr
C CTC ACC

J1	8	Arg Asp Pro Thr Val Pro Leu Ala Arg CGT GAC CCC ACC GTC CCC CTT GCG CGG
J1	35	Ala Ala Trp Glu Thr Ala Arg His Thr GCT GCG TGG GAG ACA GCT AGA CAC ACT C g Thr(=)
J1	62	Pro Val Asn Ser Trp Leu Gly Asn Ile CCA GTC AAC TCC TGG CTA GGC AAC ATC
J1	89	Ile Met Tyr Ala Pro Thr Leu Trp Ala ATC ATG TAT GCG CCC ACT TTG TGG GCA T g Ile(=)
J1	116	Arg Met Ile Leu Met Thr His Phe Phe AGG ATG ATT CTG ATG ACT CAC TTC TTC
J1	143	Ser Ile Leu Leu Ala Gln Glu Gln Leu TCC ATC CTT CTA GCC CAG GAG CAA CTT
J1	170	Glu Lys Ala Leu Asp Cys Gln Ile Tyr GAA AAA GCC CTG GAT TGT CAA ATC TAC
J1	197	Gly Ala Cys Tyr Ser Ile Glu Pro Leu GGG GCC TGT TAC TCC ATT GAG CCA CTT

FIG. 5-1

J1	224	Asp Leu Pro Gln Ile Ile Glu Arg Leu GAC CTA CCT CAG ATC ATT GAA CGA CTC
J1	251	His Gly Leu Ser Ala Phe Ser Leu His CAT GGT CTT AGC GCA TTT TCA CTC CAT
J1	278	Ser Tyr Ser Pro Gly Glu Ile Asn Arg AGT TAC TCT CCA GGT GAG ATC AAT AGG
J1	305	Val Ala Ser Cys Leu Arg Lys Leu Gly GTG GCT TCA TGC CTC AGG AAG CTT GGG
J1	332	Val Pro Pro Leu Arg Val Trp Arg His GTA CCA CCC TTG CGA GTC TGG AGA CAT
J1	359	Arg Ala Arg Ser Val Arg Ala Lys Leu CGG GCC AGA AGT GTC CGC GCT AAG CTA
J1	386	Leu Ser Gln Gly Gly Arg Ala Ala Thr CTG TCC CAA GGG GGG AGG GCC GCC ACT
		G g Gln(=)
J1	413	Lys Gly Lys Tyr Leu TGT GGC AAG TAC CTC

FIG. 5-2

J7 HCV1	1	AGCCGAGTAGTGTGGGTGCGAAAGGCCTTGTGGT
J7 HCV1	37	ACTGCCTGATAGGGTGCTTGCAGTGCCCCGGGAGG
J7 HCV1	73	<div style="display: flex; justify-content: space-between;"> <div>TCTCGTAGACCGTGCATC</div> <div> Met Ser Thr Asn ATG AGC ACA AAT C G </div> </div>
J7 HCV1	103	<div style="display: flex; justify-content: space-between;"> <div>Pro Lys Pro Gln Arg Lys Thr Lys Arg</div> </div> <div style="display: flex; justify-content: space-between;"> <div>CCT AAA CCC CAA AGA AAA ACC AAA CGT</div> </div> <div style="display: flex; justify-content: space-between;"> <div>T</div> <div>A</div> <div>A</div> </div> <div style="display: flex; justify-content: space-between;"> <div>Lys</div> <div>Asn</div> </div> <div style="display: flex; justify-content: space-between;"> <div>***</div> </div>
J7 HCV1	130	<div style="display: flex; justify-content: space-between;"> <div>Asn Thr Asn Arg Arg Pro Gln Asp Val</div> </div> <div style="display: flex; justify-content: space-between;"> <div>AAC ACC AAC CGT CGC CCA CAG GAC GTT</div> </div> <div style="display: flex; justify-content: space-between;"> <div></div> <div>C</div> </div>
J7 HCV1	157	<div style="display: flex; justify-content: space-between;"> <div>Lys Phe Pro Gly Gly Gly Gln Ile Val</div> </div> <div style="display: flex; justify-content: space-between;"> <div>AAG TTC CCG GGC GGT GGT CAG ATC GTC</div> </div> <div style="display: flex; justify-content: space-between;"> <div>T</div> <div>C</div> <div>T</div> </div>
J7 HCV1	184	<div style="display: flex; justify-content: space-between;"> <div>Gly Gly Val Tyr Leu Leu Pro Arg Arg</div> </div> <div style="display: flex; justify-content: space-between;"> <div>GGT GGA GTT TAC TTG TTG CCG CGC AGG</div> </div>
J7 HCV1	211	<div style="display: flex; justify-content: space-between;"> <div>Gly Pro Arg Leu Gly Val Arg Ala Thr</div> </div> <div style="display: flex; justify-content: space-between;"> <div>GGC CCC AGG TTG GGT GTG CGT GCG ACT</div> </div> <div style="display: flex; justify-content: space-between;"> <div>T</div> <div>A</div> <div>C</div> <div>G</div> </div>
J7 HCV1	238	<div style="display: flex; justify-content: space-between;"> <div>Arg Lys Thr Ser Glu Arg Ser Gln Pro</div> </div> <div style="display: flex; justify-content: space-between;"> <div>AGG AAG ACT TCC GAG CGG TCG CAA CCT</div> </div> <div style="display: flex; justify-content: space-between;"> <div>A</div> </div>

FIG. 6-1

J7 HCV1	265	Arg	Gly	Arg	Arg	Gln	Pro	Ile	Pro	Lys
		CGT	GGA	AGG	CGA	CAA	CCT	ATC	CCC	AAG
		A	T	A	T	G				
J7 HCV1	292	Ala	Arg	Arg	Pro	Glu	Gly	Arg	Thr	Trp
		GCT	CGC	CGG	CCC	GAG	GGC	AGG	ACC	TGG
			T							
J7 HCV1	319	Ala	Gln	Pro	Gly	Tyr	Pro	Trp	Pro	Leu
		GCT	CAG	CCT	GGG	TAT	CCT	TGG	CCC	CTC
				C		C				
J7 HCV1	346	Tyr	Gly	Asn	Glu	Gly	Leu	Gly	Trp	Ala
		TAT	GGC	AAT	GAG	GGC	TTG	GGG	TGG	GCA
							GC			G
							Cys			

J7 HCV1	373	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser
		GGA	TGG	CTC	CTG	TCA	CCC	CGC	GGC	TCT
						T		T		
J7 HCV1	400	Arg	Pro	Ser	Trp	Gly	Pro	Asn	Asp	Pro
		CGG	CCT	AGT	TGG	GGC	CCC	AAT	GAC	CCC
				C				CA		
								Thr		

J7 HCV1	427	Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys
		CGG	CGT	AGG	TCG	CGT	AAT	TTG	GGT	AAG
						C				
J7 HCV1	454	Val	Ile	Asp	Thr	Leu	Thr	Cys	Gly	Phe
		GTC	ATC	GAT	ACC	CTT	ACA	TGC	GGC	TTC
							G			

FIG. 6-2

J7	481	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Leu
HCV1		GCC	GAC	CTC	ATG	GGG	TAC	ATT	CCG	CTT
								A		C
J7	508	Val	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala
HCV1		GTC	GGC	GCC	CCC	TTA	GGG	GGC	GCT	GCC
					T C T		A			
J7	535	Arg	Ala	Leu	Ala	His	Gly			
HCV1		AGG	GCC	CTG	GCA	CAT	GGT			
					G		C			

FIG. 6-3

J1	1	T	CCG	CTC	GTC	GGC	GCC	CCC	TTA	GGG	GGC
		A						T	C	T	A
			Pro	Leu	Val	Gly	Ala	Pro	Leu	Gly	Gly
J1	29		GCT	GCC	AGG	GCC	CTG	GCA	CAT	GGT	GTC
								G		C	
			Ala	Ala	Arg	Ala	Leu	Ala	His	Gly	Val
J1	56		CGG	GTT	CTG	GAG	GAC	GGC	GTG	AAC	TAT
						A					
			Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr
J1	83		GCA	ACA	GGG	AAT	TTG	CCC	GGT	TGC	TCT
						C	C	T	T		
			Ala	Thr	Gly	Asn	Leu	Pro	Gly	Cys	Ser
J1	110		TTC	TCT	ATC	TTC	CTC	TTG	GCT	CTG	CTG
							T	C	C		C
			Phe	Ser	Ile	Phe	Leu	Leu	Ala	Leu	Leu
J1	137		TCC	TGT	TTG	ACC	ATC	CCA	GCT	TCC	GCT
			T	C		T	G	G	C	G	C
							Val				
			Ser	Cys	Leu	Thr	Ile	Pro	Ala	Ser	Ala
J1	164		TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	ATA
			C	C				TCC	A	G	C
				Gln				Ser	Thr		Leu
			***					***			
			Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Ile
J1	191		TAC	CAT	GTC	ACA	AAC	GAC	TGC	TCC	AAC
				C		C	T	T		C	T
									Pro		

FIG. 7-1

J1	218	Ser Ser Ile Val Tyr Glu Ala Ala Asp TCA AGC ATT GTG TAT GAG GCG GCG GAC G T C C T
J1	245	Val Ile Met His Ala Pro Gly Cys Val GTG ATC ATG CAT GCC CCC GGG TGC GTG CC C C A T G C Ala Leu Thr
J1	272	Pro Cys Val Arg Glu Asn Asn Ser Ser CCC TGC GTT CGG GAG AAC AAT TCC TCC T T GG C G G Gly Ala ***
J1	299	Arg Cys Trp Val Ala Leu Thr Pro Thr CGT TGC TGG GTA GCG CTC ACT CCC ACG A G T G A G C T Met
J1	326	Leu Ala Ala Arg Asn Ala Ser Val Pro CTC GCG GCC AGG AAT GCC AGC GTC CCC G G C A G G AA C Val Thr Asp Gly Lys Leu *** ***
J1	353	Thr Thr Thr Leu Arg Arg His Val Asp ACT ACG ACA TTA CGA CGC CAC GTC GAC G G CAG C T T A T Ala Gln Ile ***
J1	380	Leu Leu Val Gly Thr Ala Ala Phe Cys TTG CTC GTT GGG ACG GCT GCT TTC TGC C T C GC C A C C T Ser Thr Leu

FIG. 7-2

J1	407	Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys
		TCC	GCT	ATG	TAC	GTG	GGG	GAT	CTC	TGC
		G	C	C	C			C	A	
				Leu						
J1	434	Gly	Ser	Val	Phe	Leu	Ile	Ser	Gln	Leu
		GGA	TCT	GTT	TTC	CTC	ATC	TCC	CAG	CTG
		G		C	T	T	G	GG	A	
						Val	Gly			
J1	461	Phe	Thr	Phe	Ser	Pro	Arg	Arg	His	Glu
		TTC	ACC	TTC	TCG	CCT	CGC	CGG	CAT	GAG
					T	C	A	G	C	TG
										Trp

J1	488	Thr	Val	Gln	Asp	Cys	Asn	Cys	Ser	Ile
		ACA	GTA	CAG	GAC	TGC	AAC	TGC	TCA	ATC
		G	ACG	A	GT		T		T	
			Thr		Gly					
			***		***					
J1	515	Tyr	Pro	Gly	His	Val	Ser	Gly	His	Arg
		TAT	CCC	GGC	CAC	GTA	TCA	GGC	CAT	CGC
					T	A	A	G	T	C
					Ile	Thr				
J1	542	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp
		ATG	GCT	TGG	GAT	ATG	ATG	ATG	AAC	TGG
			A							
J1	569	Ser	Pro	Thr	Ala					
		TCG	CCC	ACG	GCA					
		C	T		A	G				
					Thr					

FIG. 7-3

			Asn	Trp	Ser	Pro	Thr	Ala	
J1	1	AAC	TGG	TCG	CCC	ACG	GCA		
HCV1				C	T		A	G	
							Thr		
		Ala	Leu	Val	Val	Ser	Gln	Leu	Leu
J1	19	GCC	TTA	GTG	GTG	TCG	CAG	TTA	CTC
HCV1		G	G	A	A	G	T	C	G
				Met	Ala				
		Ile	Pro	Gln	Ala	Val	Met	Asp	Met
J1	46	ATC	CCA	CAA	GCT	GTC	ATG	GAC	ATG
HCV1					C	A	T		
						Ile	Leu		Ile
		Ala	Gly	Ala	His	Trp	Gly	Val	Leu
J1	73	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTA
HCV1		T	T	T					G
		Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val
J1	100	GGC	CTT	GCC	TAC	TAT	TCC	ATG	GTG
HCV1			A	A	G	T	TC		
			Ile			Phe			
		Asn	Trp	Ala	Lys	Val	Leu	Ile	Val
J1	127	AAC	TGG	GCT	AAG	GTT	TTG	ATT	GTG
HCV1				G		C	C	G	A
								Val	Leu
		Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly
J1	154	CTA	CTC	TTT	GCC	GGC	GTT	GAC	GGG
HCV1		G	A					C	
									C
									Ala
									Glu

		Thr	Arg	Val	Thr	Gly	Gly	Val	Gln
J1	181	ACC	CGC	GTG	ACG	GGG	GGG	GTG	CAA
HCV1			A	C	C		A	AGT	GCC
			His					Ser	Ala
								***	***

FIG. 8-1

J1	208	His Val Thr Ser Thr Leu Thr Ser Leu
HCV1		CAC GTC ACC TCT ACA CTC ACG TCC CTC
		ACT GTG GG T T GTT AG
		Thr Val Gly Phe Val
		*** *** ***
J1	235	Phe Arg Pro Gly Ala Ser Gln Lys Ile
HCV1		TTT AGA CCT GGG GCG TCC CAG AAA ATT
		C C GC A C C AAG C G C
		Leu Ala Lys Asn Val
		*** *** ***
J1	262	Gln Leu Val Asn Thr Asn Gly Ser Trp
HCV1		CAG CTT GTA AAC ACC AAT GGC AGT TGG
		G A C C
		Ile
J1	289	His Ile Asn Arg Thr Ala Leu Asn Cys
HCV1		CAT ATC AAC AGG ACT GCC CTG AAC TGC
		C C T C G
		Leu Ser

J1	316	Asn Asp Ser Leu Gln Thr Gly Phe Leu
HCV1		AAT GAC TCC CTC CAA ACT GGG TTC CTT
		T AG A C C C GG T G
		Asn Trp
J1	343	Ala Ala Leu
HCV1		GCC GCG CTG
		A G T
		Gly

FIG. 8-2

J1	1							Ser	Val	Ile	
HCV1		ggctataccggcgacttcga						C TCA	GTG	ATC	
								G		A	
J1	11		Asp	Cys	Asn	Thr	Cys	Val	Thr	Gln	Thr
HCV1			GAC	TGT	AAC	ACA	TGT	GTC	ACT	CAG	ACG
				C	T	G			C		A
J1	38		Val	Asp	Phe	Ser	Leu	Asp	Pro	Thr	Phe
HCV1			GTC	GAT	TTC	AGC	TTG	GAT	CCC	ACC	TTC
							C T	C	T		
J1	65		Thr	Ile	Glu	Thr	Thr	Thr	Val	Pro	Gln
HCV1			ACC	ATC	GAG	ACG	ACG	ACC	GTG	CCC	CAA
				T		A	TC	G	C	C	G
							<u>Ile</u>				
J1	92		Asp	Ala	Val	Ser	Arg	Thr	Gln	Arg	Arg
HCV1			GAT	GCG	GTT	TCG	CGC	ACG	CAG	CGG	CGA
				T	C	C		T	A	T	G
J1	119		Gly	Arg	Thr	Gly	Arg	Gly	Arg	Arg	Gly
HCV1			GGT	AGG	ACT	GGC	AGG	GGC	AGG	AGA	GGC
			C					G	A	CC	
									Lys	<u>Pro</u>	
J1	146		Ile	Tyr	Arg	Phe	Val	Thr	Pro	Gly	Glu
HCV1			ATC	TAT	AGG	TTT	GTG	ACT	CCA	GGA	GAA
				C	A			G A	G	G	G
							<u>Ala</u>				
J1	173		Arg	Pro	Ser	Ala	Met	Phe	Asp	Ser	Ser
HCV1			CGG	CCC	TCG	GCG	ATG	TTC	GAT	TCT	TCG
			C		C	GC			C	G	C
						<u>Gly</u>					

FIG. 9-1

J1	200	Val	Leu	Cys	Glu	Cys	Tyr	Asp	Ala	Gly
HCV1		GTC	CTA	TGT	GAG	TGT	TAT	GAC	GCG	GGC
			C			C			A	
J1	227	Cys	Ala	Trp	Tyr	Glu	Leu	Thr	Pro	Ala
HCV1		TGT	GCT	TGG	TAT	GAG	CTC	ACG	CCC	GCT
										C
J1	254	Glu	Thr	Ser	Val	Arg	Leu	Arg	Ala	Tyr
HCV1		GAG	ACC	TCG	GTT	AGG	TTG	CGG	GCT	TAC
			T	A	A		C	A	A	G
				Thr						
J1	281	Leu	Asn	Thr	Pro	Gly	Leu	Pro	Val	Cys
HCV1		CTA	AAT	ACA	CCA	GGG	TTG	CCC	GTC	TGC
		A	G	C	C	G	C	T		G
		Met								
J1	308	Gln	Asp	His	Leu	Glu	Phe	Trp	Glu	Ser
HCV1		CAG	GAC	CAT	CTG	GAG	TTC	TGG	GAG	AGC
					T	A	T			G
										Gly
J1	335	Val	Phe	Thr	Gly	Leu	Thr	His	Ile	Asp
HCV1		GTC	TTC	ACA	GGC	CTC	ACC	CAC	ATA	GAC
			T				T	T		T
J1	362	Ala	His	Phe	Leu	Ser	Gln	Thr	Lys	Gln
HCV1		GCC	CAC	TTC	TTG	TCC	CAG	ACT	AAG	CAG
				T	C	A		A		
J1	389	Ala	Gly	Asp	Asn	Phe	Pro	Tyr	Leu	Val
HCV1		GCA	GGA	GAC	AAC	TTC	CCC	TAC	CTG	GTA
		AGT	G	G		C	T	T		
		Ser		Glu		Leu				

FIG. 9-2

J1	416	Ala Tyr Gln Ala Thr Val Cys Ala Arg
HCV1		GCA TAC CAA GCC ACA GTG TGC GCC AGG
		G C T
J1	443	Ala Lys Ala Pro Pro Pro Ser Trp Asp
HCV1		GCT AAG GCT CCA CCT CCA TCG TGG GAT
		C A C T C C
		Gln
J1	470	Gln Met Trp Lys Cys Leu Ile Arg Leu
HCV1		CAA ATG TGG AAG TGT CTC ATA CGG CTA
		G T G T C C
J1	497	Lys Pro Thr Leu His Gly Pro Thr Pro
HCV1		AAG CCT ACG CTG CAC GGG CCA ACG CCC
		C C C T A
J1	524	Leu Leu Tyr Arg Leu Gly Ala Val Gln
HCV1		CTG CTG TAT AGG CTA GGA GCC GTC CAG
		A C A G C T T
J1	551	Asn Glu Val Thr Leu Thr His Pro Ile
HCV1		AAT GAG GTC ACC CTC ACA CAC CCT ATA
		A A G G A G C
		Ile Val
J1	578	Thr Lys
HCV1		ACC AAA
		tacatcatgacatgcatgtc

FIG. 9-3

		J1		1		Leu Thr		C CTC ACC		
		HCV1								
J1	8	Arg	Asp	Pro	Thr	Val	Pro	Leu	Ala	Arg
HCV1		CGT	GAC	CCC	ACC	GTC	CCC	CTT	GCG	CGG
				T	A	AC		C		A A
						Thr				

J1	35	Ala	Ala	Trp	Glu	Thr	Ala	Arg	His	Thr
HCV1		GCT	GCG	TGG	GAG	ACA	GCT	AGA	CAC	ACT
							A			
J1	62	Pro	Val	Asn	Ser	Trp	Leu	Gly	Asn	Ile
HCV1		CCA	GTC	AAC	TCC	TGG	CTA	GGC	AAC	ATC
				T						A
J1	89	Ile	Met	Tyr	Ala	Pro	Thr	Leu	Trp	Ala
HCV1		ATC	ATG	TAT	GCG	CCC	ACT	TTG	TGG	GCA
				T	C		A	C		G
				Phe						
J1	116	Arg	Met	Ile	Leu	Met	Thr	His	Phe	Phe
HCV1		AGG	ATG	ATT	CTG	ATG	ACT	CAC	TTC	TTC
				A			C	T		T
J1	143	Ser	Ile	Leu	Leu	Ala	Gln	Glu	Gln	Leu
HCV1		TCC	ATC	CTT	CTA	GCC	CAG	GAG	CAA	CTT
		AG	G		A		AG	C	G	
			Val		Ile		Arg	Asp		

J1	170	Glu	Lys	Ala	Leu	Asp	Cys	Gln	Ile	Tyr
HCV1		GAA	AAA	GCC	CTG	GAT	TGT	CAA	ATC	TAC
			C G		C		C	G G		
			Gln					Glu		
			***					***		

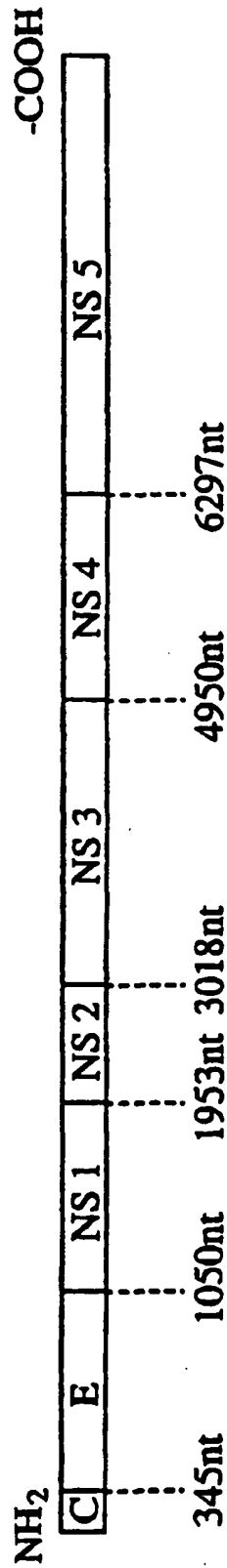
FIG. 10-1

J1	197	Gly	Ala	Cys	Tyr	Ser	Ile	Glu	Pro	Leu
HCV1		GGG	GCC	TGT	TAC	TCC	ATT	GAG	CCA	CTT
				C			A	A		
J1	224	Asp	Leu	Pro	Gln	Ile	Ile	Glu	Arg	Leu
HCV1		GAC	CTA	CCT	CAG	ATC	ATT	GAA	CGA	CTC
		T			CA			C	A	
					Pro			Gln		
					***			***		
J1	251	His	Gly	Leu	Ser	Ala	Phe	Ser	Leu	His
HCV1		CAT	GGT	CTT	AGC	GCA	TTT	TCA	CTC	CAT
			C	C						C
J1	278	Ser	Tyr	Ser	Pro	Gly	Glu	Ile	Asn	Arg
HCV1		AGT	TAC	TCT	CCA	GGT	GAG	ATC	AAT	AGG
							A	T		
J1	305	Val	Ala	Ser	Cys	Leu	Arg	Lys	Leu	Gly
HCV1		GTG	GCT	TCA	TGC	CTC	AGG	AAG	CTT	GGG
			C	G			A	A		
				Ala						
J1	332	Val	Pro	Pro	Leu	Arg	Val	Trp	Arg	His
HCV1		GTA	CCA	CCC	TTG	CGA	GTC	TGG	AGA	CAT
			G				CT			C
							Ala			

J1	359	Arg	Ala	Arg	Ser	Val	Arg	Ala	Lys	Leu
HCV1		CGG	GCC	AGA	AGT	GTC	CGC	GCT	AAG	CTA
				C	C				G	T
									Arg	
J1	386	Leu	Ser	Gln	Gly	Gly	Arg	Ala	Ala	Thr
HCV1		CTG	TCC	CAA	GGG	GGG	AGG	GCC	GCC	ACT
			G	AG	A	C		T		TA
			Ala	Arg						Ile
				***						***

FIG. 10-2

J1 Lys Gly Lys Tyr Leu
 HCV1 413 TGT GGC AAG TAC CTC
 FIG. 10-3



The nucleotide numbers are approximate
 FIG. 11

-267 GCGTCTAGCCATGGCGTTAGTATGAGTGTCTGTCAGCCTCCAGG
 CGCAGATCGGTACCGCAATCATACTCACAGCACGTCTGGAGGTCC

 -223 ACCCCCCCTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGA
 TGGGGGGGAGGGCCCTCTCGGTATCACCAGACGCCTTGGCCACT

 -179 GTACACCGGAATTGCCAGGACGACCGGGTCCTTTCTTGGATCAA
 CATGTGGCCTTAACGGTCTCTGCTGGCCCAGGAAAGAACCTAGTT

 -135 CCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCCGCAAGACTGC
 GGGCGAGTTACGGACCTCTAAACCCGCACGGGGGCGTTCTGACG

 -91 TAGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCT
 ATCGGCTCATCACAACCCAGCGCTTCCGGAACACCATGACGGA

 -47 GATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGC
 CTATCCACGAACGCTCACGGGGCCCTCCAGAGCATCTGGCACG

 -3 ACC -1
 TGG

1	Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Lys	Lys	Asn
	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AAA	AAA	AAC
	TAC	TCG	TGC	TTA	GGA	TTT	GGA	GTT	TTT	TTT	TTG
34	Lys	Arg	Asn	Thr	Asn	Arg	Arg	Pro	Gln	Asp	Val
	AAA	CGT	AAC	ACC	AAC	CGT	CGC	CCA	CAG	GAC	GTC
	TTT	GCA	TTG	TGG	TTG	GCA	GCG	GGT	GTC	CTG	CAG
67	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile	Val	Gly	Gly
	AAG	TTC	CCG	GGT	GGC	GGT	CAG	ATC	GTT	GGT	GGA
	TTC	AAG	GGC	CCA	CCG	CCA	GTC	TAG	CAA	CCA	CCT

FIG. 12-1

100	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro	Arg	Leu
	GTT	TAC	TTG	TTG	CCG	CGC	AGG	GGC	CCT	AGA	TTG
	CAA	ATG	AAC	AAC	GGC	GCG	TCC	CCG	GGA	TCT	AAC
133	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg
	GGT	GTG	CGC	GCG	ACG	AGA	AAG	ACT	TCC	GAG	CGG
	CCA	CAC	GCG	CGC	TGC	TCT	TTC	TGA	AGG	CTC	GCC
166	Ser	Gln	Pro	Arg	Gly	Arg	Arg	Gln	Pro	Ile	Pro
	TCG	CAA	CCT	CGA	GGT	AGA	CGT	CAG	CCT	ATC	CCC
	AGC	GTT	GGA	GCT	CCA	TCT	GCA	GTC	GGA	TAG	GGG
199	Lys	Ala	Arg	Arg	Pro	Glu	Gly	Arg	Thr	Trp	Ala
	AAG	GCT	CGT	CGG	CCC	GAG	GGC	AGG	ACC	TGG	GCT
	TTC	CGA	GCA	GCC	GGG	CTC	CCG	TCC	TGG	ACC	CGA
232	Gln	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn
	CAG	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT
	GTC	GGG	CCC	ATG	GGA	ACC	GGG	GAG	ATA	CCG	TTA
265	Glu	Gly	Cys	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Ser
	GAG	GGC	TGC	GGG	TGG	GCG	GGA	TGG	CTC	CTG	TCT
	CTC	CCG	ACG	CCC	ACC	CGC	CCT	ACC	GAG	GAC	AGA
298	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	Gly	Pro	Thr
	CCC	CGT	GGC	TCT	CGG	CCT	AGC	TGG	GGC	CCC	ACA
	GGG	GCA	CCG	AGA	GCC	GGA	TCG	ACC	CCG	GGG	TGT
331	Asp	Pro	Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys
	GAC	CCC	CGG	CGT	AGG	TCG	CGC	AAT	TTG	GGT	AAG
	CTG	GGG	GCC	GCA	TCC	AGC	GCG	TTA	AAC	CCA	TTC
364	Val	Ile	Asp	Thr	Leu	Thr	Cys	Gly	Phe	Ala	Asp
	GTC	ATC	GAT	ACC	CTT	ACG	TGC	GGC	TTC	GCC	GAC
	CAG	TAG	CTA	TGG	GAA	TGC	ACG	CCG	AAG	CGG	CTG

FIG. 12-2

397	Leu	Met	Gly	Tyr	Ile	Pro	Leu	Val	Gly	Ala	Pro
	CTC	ATG	GGG	TAC	ATA	CCG	CTC	GTC	GGC	GCC	CCT
	GAG	TAC	CCC	ATG	TAT	GGC	GAG	CAG	CCG	CGG	GGA
430	Leu	Gly	Gly	Ala	Ala	Arg	Ala	Leu	Ala	His	Gly
	CTT	GGA	GGC	GCT	GCC	AGG	GCC	CTG	GCG	CAT	GGC
	GAA	CCT	CCG	CGA	CGG	TCC	CGG	GAC	CGC	GTA	CCG
463	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr	Ala
	GTC	CGG	GTT	CTG	GAA	GAC	GGC	GTG	AAC	TAT	GCA
	CAG	GCC	CAA	GAC	CTT	CTG	CCG	CAC	TTG	ATA	CGT
496	Thr	Gly	Asn	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile
	ACA	GGG	AAC	CTT	CCT	GGT	TGC	TCT	TTC	TCT	ATC
	TGT	CCC	TTG	GAA	GGA	CCA	ACG	AGA	AAG	AGA	TAG
529	Phe	Leu	Leu	Ala	Leu	Leu	Ser	Cys	Leu	Thr	Val
	TTC	CTT	CTG	GCC	CTG	CTC	TCT	TGC	TTG	ACT	GTG
	AAG	GAA	GAC	CGG	GAC	GAG	AGA	ACG	AAC	TGA	CAC
562	Pro	Ala	Ser	Ala	Tyr	Gln	Val	Arg	Asn	Ser	Thr
	CCC	GCT	TCG	GCC	TAC	CAA	GTG	CGC	AAC	TCC	ACG
	GGG	CGA	AGC	CGG	ATG	GTT	CAC	GCG	TTG	AGG	TGC
595	Gly	Leu	Tyr	His	Val	Thr	Asn	Asp	Cys	Pro	Asn
	GGG	CTT	TAC	CAC	GTC	ACC	AAT	GAT	TGC	CCT	AAC
	CCC	GAA	ATG	GTG	CAG	TGG	TTA	CTA	ACG	GGA	TTG
628	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Ala	Asp	Ala	Ile
	TCG	AGT	ATT	GTG	TAC	GAG	GCG	GCC	GAT	GCC	ATC
	AGC	TCA	TAA	CAC	ATG	CTC	CGC	CGG	CTA	CGG	TAG
661	Leu	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg
	CTG	CAC	ACT	CCG	GGG	TGC	GTC	CCT	TGC	GTT	CGT
	GAC	GTG	TGA	GGC	CCC	ACG	CAG	GGA	ACG	CAA	GCA

FIG. 12-3

694	Glu Gly Asn Ala Ser Arg Cys Trp Val Ala Met
	GAG GGC AAC GCC TCG AGG TGT TGG GTG GCG ATG
	CTC CCG TTG CCG AGC TCC ACA ACC CAC CGC TAC
727	Thr Pro Thr Val Ala Thr Arg Asp Gly Lys Leu
	ACC CCT ACG GTG GCC ACC AGG GAT GGC AAA CTC
	TGG GGA TGC CAC CGG TGG TCC CTA CCG TTT GAG
760	Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu
	CCC GCG ACG CAG CTT CGA CGT CAC ATC GAT CTG
	GGG CGC TGC GTC GAA GCT GCA GTG TAG CTA GAC
793	Leu Val Gly Ser Ala Thr Leu Cys Ser Ala Leu
	CTT GTC GGG AGC GCC ACC CTC TGT TCG GCC CTC
	GAA CAG CCC TCG CGG TGG GAG ACA AGC CGG GAG
826	Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu
	TAC GTG GGG GAC CTA TGC GGG TCT GTC TTT CTT
	ATG CAC CCC CTG GAT ACG CCC AGA CAG AAA GAA
859	Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg
	GTC GGC CAA CTG TTC ACC TTC TCT CCC AGG CGC
	CAG CCG GTT GAC AAG TGG AAG AGA GGG TCC GCG
892	His Trp Thr Thr Gln Gly Cys Asn Cys Ser Ile
	CAC TGG ACG ACG CAA GGT TGC AAT TGC TCT ATC
	GTG ACC TGC TGC GTT CCA ACG TTA ACG AGA TAG
925	Tyr Pro Gly His Ile Thr Gly His Arg Met Ala
	TAT CCC GGC CAT ATA ACG GGT CAC CGC ATG GCA
	ATA GGG CCG GTA TAT TGC CCA GTG GCG TAC CGT
958	Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr
	TGG GAT ATG ATG ATG AAC TGG TCC CCT ACG ACG
	ACC CTA TAC TAC TAC TTG ACC AGG GGA TGC TGC

FIG. 12-4

991	Ala	Leu	Val	Met	Ala	Gln	Leu	Leu	Arg	Ile	Pro
	GCG	TTG	GTA	ATG	GCT	CAG	CTG	CTC	CGG	ATC	CCA
	CGC	AAC	CAT	TAC	CGA	GTC	GAC	GAG	GCC	TAG	GGT
1024	Gln	Ala	Ile	Leu	Asp	Met	Ile	Ala	Gly	Ala	His
	CAA	GCC	ATC	TTG	GAC	ATG	ATC	GCT	GGT	GCT	CAC
	GTT	CGG	TAG	AAC	CTG	TAC	TAG	CGA	CCA	CGA	GTG
1057	Trp	Gly	Val	Leu	Ala	Gly	Ile	Ala	Tyr	Phe	Ser
	TGG	GGA	GTC	CTG	GCG	GGC	ATA	GCG	TAT	TTC	TCC
	ACC	CCT	CAG	GAC	CGC	CCG	TAT	CGC	ATA	AAG	AGG
1090	Met	Val	Gly	Asn	Trp	Ala	Lys	Val	Leu	Val	Val
	ATG	GTG	GGG	AAC	TGG	GCG	AAG	GTC	CTG	GTA	GTG
	TAC	CAC	CCC	TTG	ACC	CGC	TTC	CAG	GAC	CAT	CAC
1123	Leu	Leu	Leu	Phe	Ala	Gly	Val	Asp	Ala	Glu	Thr
	CTG	CTG	CTA	TTT	GCC	GGC	GTC	GAC	GCG	GAA	ACC
	GAC	GAC	GAT	AAA	CGG	CCG	CAG	CTG	CGC	CTT	TGG
1156	His	Val	Thr	Gly	Gly	Ser	Ala	Gly	His	Thr	Val
	CAC	GTC	ACC	GGG	GGA	AGT	GCC	GGC	CAC	ACT	GTG
	GTG	CAG	TGG	CCC	CCT	TCA	CGG	CCG	GTG	TGA	CAC
1189	Ser	Gly	Phe	Val	Ser	Leu	Leu	Ala	Pro	Gly	Ala
	TCT	GGA	TTT	GTT	AGC	CTC	CTC	GCA	CCA	GGC	GCC
	AGA	CCT	AAA	CAA	TCG	GAG	GAG	CGT	GGT	CCG	CGG
1222	Lys	Gln	Asn	Val	Gln	Leu	Ile	Asn	Thr	Asn	Gly
	AAG	CAG	AAC	GTC	CAG	CTG	ATC	AAC	ACC	AAC	GGC
	TTC	GTC	TTG	CAG	GTC	GAC	TAG	TTG	TGG	TTG	CCG
1255	Ser	Trp	His	Leu	Asn	Ser	Thr	Ala	Leu	Asn	Cys
	AGT	TGG	CAC	CTC	AAT	AGC	ACG	GCC	CTG	AAC	TGC
	TCA	ACC	GTG	GAG	TTA	TCG	TGC	CGG	GAC	TTG	ACG

FIG. 12-5

1288	Asn Asp Ser Leu Asn Thr Gly Trp Leu Ala Gly
	AAT GAT AGC CTC AAC ACC GGC TGG TTG GCA GGG
	TTA CTA TCG GAG TTG TGG CCG ACC AAC CGT CCC
1321	Leu Phe Tyr His His Lys Phe Asn Ser Ser Gly
	TTT TCT ATC ACC ACA AGT TCA ACT CTT CAG GCT
	GAA AAG ATA GTG GTG TTC AAG TTG AGA AGT CCG
1354	Cys Pro Glu Arg Leu Ala Ser Cys Arg Pro Leu
	GTC CTG AGA GGC TAG CCA GCT GCC GAC CCC CTT
	ACA GGA CTC TCC GAT CGG TCG ACG GCT GGG GAA
1387	Thr Asp Phe Asp Gln Gly Trp Gly Pro Ile Ser
	ACC GAT TTT GAC CAG GGC TGG GGC CCT ATC AGT
	TGG CTA AAA CTG GTC CCG ACC CCG GGA TAG TCA
1420	Tyr Ala Asn Gly Ser Gly Pro Asp Gln Arg Pro
	TAT GCC AAC GGA AGC GGC CCC GAC CAG CGC CCC
	ATA CGG TTG CCT TCG CCG GGG CTG GTC GCG GGG
1453	Tyr Cys Trp His Tyr Pro Pro Lys Pro Cys Gly
	TAC TGC TGG CAC TAC CCC CCA AAA CCT TGC GGT
	ATG ACG ACC GTG ATG GGG GGT TTT GGA ACG CCA
1486	Ile Val Pro Ala Lys Ser Val Cys Gly Pro Val
	ATT GTG CCC GCG AAG AGT GTG TGT GGT CCG GTA
	TAA CAC GGG CGC TTC TCA CAC ACA CCA GGC CAT
1519	Tyr Cys Phe Thr Pro Ser Pro Val Val Val Gly
	TAT TGC TTC ACT CCC AGC CCC GTG GTG GTG GGA
	ATA ACG AAG TGA GGG TCG GGG CAC CAC CAC CCT
1552	Thr Thr Asp Arg Ser Gly Ala Pro Thr Tyr Ser
	ACG ACC GAC AGG TCG GGC GCG CCC ACC TAC AGC
	TGC TGG CTG TCC AGC CCG CGC GGG TGG ATG TCG

FIG. 12-6

1585	Trp	Gly	Glu	Asn	Asp	Thr	Asp	Val	Phe	Val	Leu
	TGG	GGT	GAA	AAT	GAT	ACG	GAC	GTC	TTC	GTC	CTT
	ACC	CCA	CTT	TTA	CTA	TGC	CTG	CAG	AAG	CAG	GAA
1618	Asn	Asn	Thr	Arg	Pro	Pro	Leu	Gly	Asn	Trp	Phe
	AAC	AAT	ACC	AGG	CCA	CCG	CTG	GGC	AAT	TGG	TTC
	TTG	TTA	TGG	TCC	GGT	GGC	GAC	CCG	TTA	ACC	AAG
1651	Gly	Cys	Thr	Trp	Met	Asn	Ser	Thr	Gly	Phe	Thr
	GGT	TGT	ACC	TGG	ATG	AAC	TCA	ACT	GGA	TTC	ACC
	CCA	ACA	TGG	ACC	TAC	TTG	AGT	TGA	CCT	AAG	TGG
1684	Lys	Val	Cys	Gly	Ala	Pro	Pro	Cys	Val	Ile	Gly
	AAA	GTG	TGC	GGA	GCG	CCT	CCT	TGT	GTC	ATC	GGA
	TTT	CAC	ACG	CCT	CGC	GGA	GGA	ACA	CAG	TAG	CCT
1717	Gly	Ala	Gly	Asn	Asn	Thr	Leu	His	Cys	Pro	Thr
	GGG	GCG	GGC	AAC	AAC	ACC	CTG	CAC	TGC	CCC	ACT
	CCC	CGC	CCG	TTG	TTG	TGG	GAC	GTG	ACG	GGG	TGA
1750	Asp	Cys	Phe	Arg	Lys	His	Pro	Asp	Ala	Thr	Tyr
	GAT	TGC	TTC	CGC	AAG	CAT	CCG	GAC	GCC	ACA	TAC
	CTA	ACG	AAG	GCG	TTC	GTA	GGC	CTG	CGG	TGT	ATG
1783	Ser	Arg	Cys	Gly	Ser	Gly	Pro	Trp	Ile	Thr	Pro
	TCT	CGG	TGC	GGC	TCC	GGT	CCC	TGG	ATC	ACA	CCC
	AGA	GCC	ACG	CCG	AGG	CCA	GGG	ACC	TAG	TGT	GGG
1816	Arg	Cys	Leu	Val	Asp	Tyr	Pro	Tyr	Arg	Leu	Trp
	AGG	TGC	CTG	GTC	GAC	TAC	CCG	TAT	AGG	CTT	TGG
	TCC	ACG	GAC	CAG	CTG	ATG	GGC	ATA	TCC	GAA	ACC
1849	His	Tyr	Pro	Cys	Thr	Ile	Asn	Tyr	Thr	Ile	Phe
	CAT	TAT	CCT	TGT	ACC	ATC	AAC	TAC	ACC	ATA	TTT
	GTA	ATA	GGA	ACA	TGG	TAG	TTG	ATG	TGG	TAT	AAA

FIG. 12-7

1882	Lys	Ile	Arg	Met	Tyr	Val	Gly	Gly	Val	Glu	His
	AAA	ATC	AGG	ATG	TAC	GTG	GGA	GGG	GTC	GAA	CAC
	TTT	TAG	TCC	TAC	ATG	CAC	CCT	CCC	CAG	CTT	GTG
1915	Arg	Leu	Glu	Ala	Ala	Cys	Asn	Trp	Thr	Arg	Gly
	AGG	CTG	GAA	GCT	GCC	TGC	AAC	TGG	ACG	CGG	GGC
	TCC	GAC	CTT	CGA	CGG	ACG	TTG	ACC	TGC	GCC	CCG
1948	Glu	Arg	Cys	Asp	Leu	Glu	Asp	Arg	Asp	Arg	Ser
	GAA	CGT	TGC	GAT	CTG	GAA	GAC	AGG	GAC	AGG	TCC
	CTT	GCA	ACG	CTA	GAC	CTT	CTG	TCC	CTG	TCC	AGG
1981	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Thr	Thr	Thr	Gln
	GAG	CTC	AGC	CCG	TTA	CTG	CTG	ACC	ACT	ACA	CAG
	CTC	GAG	TCG	GGC	AAT	GAC	GAC	TGG	TGA	TGT	GTG
2014	Trp	Gln	Val	Leu	Pro	Cys	Ser	Phe	Thr	Thr	Leu
	TGG	CAG	GTC	CTC	CCG	TGT	TCC	TTC	ACA	ACC	CTA
	ACC	GTC	CAG	GAG	GGC	ACA	AGG	AAG	TGT	TGG	GAT
2047	Pro	Ala	Leu	Ser	Thr	Gly	Leu	Ile	His	Leu	His
	CCA	GCC	TTG	TCC	ACC	GGC	CTC	ATC	CAC	CTC	CAC
	GGT	CGG	AAC	AGG	TGG	CCG	GAG	TAG	GTG	GAG	GTG
2080	Gln	Asn	Ile	Val	Asp	Val	Gln	Tyr	Leu	Tyr	Gly
	CAG	AAC	ATT	GTG	GAC	GTG	CAG	TAC	TTG	TAC	GGG
	GTC	TTG	TAA	CAC	CTG	CAC	GTC	ATG	AAC	ATG	CCC
2113	Val	Gly	Ser	Ser	Ile	Ala	Ser	Trp	Ala	Ile	Lys
	GTG	GGG	TCA	AGC	ATC	GCG	TCC	TGG	GCC	ATT	AAG
	CAC	CCC	AGT	TCG	TAG	CGC	AGG	ACC	CGG	TAA	TTC
2146	Trp	Glu	Tyr	Val	Val	Leu	Leu	Phe	Leu	Leu	Leu
	TGG	GAG	TAC	GTC	GTT	CTC	CTG	TTC	CTT	CTG	CTT
	ACC	CTC	ATG	CAG	CAA	GAG	GAC	AAG	GAA	GAC	GAA

FIG. 12-8

2179	Ala	Asp	Ala	Arg	Val	Cys	Ser	Cys	Leu	Trp	Met
	GCA	GAC	GCG	CGC	GTC	TGC	TCC	TGC	TTG	TGG	ATG
	CGT	CTG	CGC	GCG	CAG	ACG	AGG	ACG	AAC	ACC	TAC
2212	Met	Leu	Leu	Ile	Ser	Gln	Ala	Glu	Ala	Ala	Leu
	ATG	CTA	CTC	ATA	TCC	CAA	GCG	GAG	GCG	GCT	TTG
	TAC	GAT	GAG	TAT	AGG	GTT	CGC	CTC	CGC	CGA	AAC
2245	Glu	Asn	Leu	Val	Ile	Leu	Asn	Ala	Ala	Ser	Leu
	GAG	AAC	CTC	GTA	ATA	CTT	AAT	GCA	GCA	TCC	CTG
	CTC	TTG	GAG	CAT	TAT	GAA	TTA	CGT	CGT	AGG	GAC
2278	Ala	Gly	Thr	His	Gly	Leu	Val	Ser	Phe	Leu	Val
	GCC	GGG	ACG	CAC	GGT	CTT	GTA	TCC	TTC	CTC	GTG
	CGG	CCC	TGC	GTG	CCA	GAA	CAT	AGG	AAG	GAG	CAC
2311	Phe	Phe	Cys	Phe	Ala	Trp	Tyr	Leu	Lys	Gly	Lys
	TTC	TTC	TGC	TTT	GCA	TGG	TAT	TTG	AAG	GGT	AAG
	AAG	AAG	ACG	AAA	CGT	ACC	ATA	AAC	TTC	CCA	TTC
2344	Trp	Val	Pro	Gly	Ala	Val	Tyr	Thr	Phe	Tyr	Gly
	TGG	GTG	CCC	GGG	GCG	GTC	TAC	ACC	TTC	TAC	GGG
	ACC	CAC	GGG	CCT	CGC	CAG	ATG	TGG	AAG	ATG	CCC
2377	Met	Trp	Pro	Leu	Leu	Leu	Leu	Leu	Leu	Ala	Leu
	ATG	TGG	CCT	CTC	CTC	CTG	CTC	CTG	TTG	GCG	TTG
	TAC	ACC	GGA	GAG	GAG	GAC	GAG	GAC	AAC	CGC	AAC
2410	Pro	Gln	Arg	Ala	Tyr	Ala	Leu	Asp	Thr	Glu	Val
	CCC	CAG	CGG	GCG	TAC	GCG	CTG	GAC	ACG	GAG	GTG
	GGG	GTC	GCC	CGC	ATG	CGC	GAC	CTG	TGC	CTC	CAC
2443	Ala	Ala	Ser	Cys	Gly	Gly	Val	Val	Leu	Val	Gly
	GCC	GCG	TCG	TGT	GGC	GGT	GTT	GTT	CTC	GTC	GGG
	CGG	CGC	AGC	ACA	CCG	CCA	CAA	CAA	GAG	CAG	CCC

FIG. 12-9

2476	Leu Met Ala Leu Thr Leu Ser Pro Tyr Tyr Lys
	TTG ATG GCG CTG ACT CTG TCA CCA TAT TAC AAG
	AAC TAC CGC GAC TGA GAC AGT GGT ATA ATG TTC
2509	Arg Tyr Ile Ser Trp Cys Leu Trp Trp Leu Gln
	CGC TAT ATC AGC TGG TGC TTG TGG TGG CTT CAG
	GCG ATA TAG TCG ACC ACG AAC ACC ACC GAA GTC
2542	Tyr Phe Leu Thr Arg Val Glu Ala Gln Leu His
	TAT TTT CTG ACC AGA GTG GAA GCG CAA CTG CAC
	ATA AAA GAC TGG TCT CAC CTT CGC GTT GAC GTG
2575	Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly
	GTG TGG ATT CCC CCC CTC AAC GTC CGA GGG GGG
	CAC ACC TAA GGG GGG GAG TTG CAG GCT CCC CCC
2608	Arg Asp Ala Val Ile Leu Leu Met Cys Ala Val
	CGC GAC GCC GTC ATC TTA CTC ATG TGT GCT GTA
	GCG CTG CGG CAG TAG AAT GAG TAC ACA CGA CAT
2641	His Pro Thr Leu Val Phe Asp Ile Thr Lys Leu
	CAC CCG ACT CTG GTA TTT GAC ATC ACC AAA TTG
	GTG GGC TGA GAC CAT AAA CTG TAG TGG TTT AAC
2674	Leu Leu Ala Val Phe Gly Pro Leu Trp Ile Leu
	CTG CTG GCC GTC TTC GGA CCC CTT TGG ATT CTT
	GAC GAC CGG CAG AAG CCT GGG GAA ACC TAA GAA
2707	Gln Ala Ser Leu Leu Lys Val Pro Tyr Phe Val
	CAA GCC AGT TTG CTT AAA GTA CCC TAC TTT GTG
	GTT CGG TCA AAC GAA TTT CAT GGG ATG AAA CAC
2740	Arg Val Gln Gly Leu Leu Arg Phe Cys Ala Leu
	CGC GTC CAA GGC CTT CTC CGG TTC TGC GCG TTA
	GCG CAG GTT CCG GAA GAG GCC AAG ACG CGC AAT

FIG. 12-10

2773	Ala	Arg	Lys	Met	Ile	Gly	Gly	His	Tyr	Val	Gln
	GCG	CGG	AAG	ATG	ATC	GGA	GGC	CAT	TAC	GTG	CAA
	CGC	GCC	TTC	TAC	TAG	CCT	CCG	GTA	ATG	CAC	GTT
2806	Met	Val	Ile	Ile	Lys	Leu	Gly	Ala	Leu	Thr	Gly
	ATG	GTC	ATC	ATT	AAG	TTA	GGG	GCG	CTT	ACT	GGC
	TAC	CAG	TAG	TAA	TTC	AAT	CCC	CGC	GAA	TGA	CCG
2839	Thr	Tyr	Val	Tyr	Asn	His	Leu	Thr	Pro	Leu	Arg
	ACC	TAT	GTT	TAT	AAC	CAT	CTC	ACT	CCT	CTT	CGG
	TGG	ATA	CAA	ATA	TTG	GTA	GAG	TGA	GGA	GAA	GCC
2872	Asp	Trp	Ala	His	Asn	Gly	Leu	Arg	Asp	Leu	Ala
	GAC	TGG	GCG	CAC	AAC	GGC	TTG	CGA	GAT	CTG	GCC
	CTG	ACC	CGC	GTG	TTG	CCG	AAC	GCT	CTA	GAC	CGG
2905	Val	Ala	Val	Glu	Pro	Val	Val	Phe	Ser	Gln	Met
	GTG	GCT	GTA	GAG	CCA	GTC	GTC	TTC	TCC	CAA	ATG
	CAC	CGA	CAT	CTC	GGT	CAG	CAG	AAG	AGG	GTT	TAC
2938	Glu	Thr	Lys	Leu	Ile	Thr	Trp	Gly	Ala	Asp	Thr
	GAG	ACC	AAG	CTC	ATC	ACG	TGG	GGG	GCA	GAT	ACC
	CTC	TGG	TTC	GAG	TAG	TGC	ACC	CCC	CGT	CTA	TGG
2971	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Asn	Gly	Leu	Pro
	GCC	GCG	TGC	GGT	GAC	ATC	ATC	AAC	GGC	TTG	CCT
	CGG	CGC	ACG	CCA	CTG	TAG	TAG	TTG	CCG	AAC	GGA
3004	Val	Ser	Ala	Arg	Arg	Gly	Arg	Glu	Ile	Leu	Leu
	GTT	TCC	GCC	CGC	AGG	GGC	CGG	GAG	ATA	CTG	CTC
	CAA	AGG	CGG	GCG	TCC	CCG	GCC	CTC	TAT	GAC	GAG
3037	Gly	Pro	Ala	Asp	Gly	Met	Val	Ser	Lys	Gly	Trp
	GGG	CCA	GCC	GAT	GGA	ATG	GTC	TCC	AAG	GGG	TGG
	CCC	GGT	CGG	CTA	CCT	TAC	CAG	AGG	TTC	CCC	ACC

FIG. 12-11

3070	Arg	Leu	Leu	Ala	Pro	Ile	Thr	Ala	Tyr	Ala	Gln
	AGG	TTG	CTG	GCG	CCC	ATC	ACG	GCG	TAC	GCC	CAG
	TCC	AAC	GAC	CGC	GGG	TAG	TGC	CGC	ATG	CGG	GTC
3103	Gln	Thr	Arg	Gly	Leu	Leu	Gly	Cys	Ile	Ile	Thr
	CAG	ACA	AGG	GGC	CTC	CTA	GGG	TGC	ATA	ATC	ACC
	GTC	TGT	TCC	CCG	GAG	GAT	CCC	ACG	TAT	TAG	TGG
3136	Ser	Leu	Thr	Gly	Arg	Asp	Lys	Asn	Gln	Val	Glu
	AGC	CTA	ACT	GGC	CGG	GAC	AAA	AAC	CAA	GTG	GAG
	TCG	GAT	TGA	CCG	GCC	CTG	TTT	TTG	GTT	CAC	CTC
3169	Gly	Glu	Val	Gln	Ile	Val	Ser	Thr	Ala	Ala	Gln
	GGT	GAG	GTC	CAG	ATT	GTG	TCA	ACT	GCT	GCC	CAA
	CCA	CTC	CAG	GTC	TAA	CAC	AGT	TGA	CGA	CGG	GTT
3202	Thr	Phe	Leu	Ala	Thr	Cys	Ile	Asn	Gly	Val	Cys
	ACC	TTC	CTG	GCA	ACG	TGC	ATC	AAT	GGG	GTG	TGC
	TGG	AAG	GAC	CGT	TGC	ACG	TAG	TTA	CCC	CAC	ACG
3235	Trp	Thr	Val	Tyr	His	Gly	Ala	Gly	Thr	Arg	Thr
	TGG	ACT	GTC	TAC	CAC	GGG	GCC	GGA	ACG	AGG	ACC
	ACC	TGA	CAG	ATG	GTG	CCC	CGG	CCT	TGC	TCC	TGG
3268	Ile	Ala	Ser	Pro	Lys	Gly	Pro	Val	Ile	Gln	Met
	ATC	GCG	TCA	CCC	AAG	GGT	CCT	GTC	ATC	CAG	ATG
	TAG	CGC	AGT	GGG	TTC	CCA	GGA	CAG	TAG	GTC	TAC
3301	Tyr	Thr	Asn	Val	Asp	Gln	Asp	Leu	Val	Gly	Trp
	TAT	ACC	AAT	GTA	GAC	CAA	GAC	CTT	GTG	GGC	TGG
	ATA	TGG	TTA	CAT	CTG	GTT	CTG	GAA	CAC	CCG	ACC
3334	Pro	Ala	Pro	Gln	Gly	Ser	Arg	Ser	Leu	Thr	Pro
	CCC	GCT	CCG	CAA	GGT	AGC	CGC	TCA	TTG	ACA	CCC
	GGG	CGA	GGC	GTT	CCA	TCG	GCG	AGT	AAC	TGT	GGG

FIG. 12-12

3367	Cys	Thr	Cys	Gly	Ser	Ser	Asp	Leu	Tyr	Leu	Val
	TGC	ACT	TGC	GGC	TCC	TCG	GAC	CTT	TAC	CTG	GTC
	ACG	TGA	ACG	CCG	AGG	AGC	CTG	GAA	ATG	GAC	CAG
3400	Thr	Arg	His	Ala	Asp	Val	Ile	Pro	Val	Arg	Arg
	ACG	AGG	CAC	GCC	GAT	GTC	ATT	CCC	GTG	CGC	CGG
	TGC	TCC	GTG	CGG	CTA	CAG	TAA	GGG	CAC	GCG	GCC
3433	Arg	Gly	Asp	Ser	Arg	Gly	Ser	Leu	Leu	Ser	Pro
	CGG	GGT	GAT	AGC	AGG	GGC	AGC	CTG	CTG	TCG	CCC
	GCC	CCA	CTA	TCG	TCC	CCG	TCG	GAC	GAC	AGC	GGG
3466	Arg	Pro	Ile	Ser	Tyr	Leu	Lys	Gly	Ser	Ser	Gly
	CGG	CCC	ATT	TCC	TAC	TTG	AAA	GGC	TCC	TCG	GGG
	GCC	GGG	TAA	AGG	ATG	AAC	TTT	CCG	AGG	AGC	CCC
3499	Gly	Pro	Leu	Leu	Cys	Pro	Ala	Gly	His	Ala	Val
	GGT	CCG	CTG	TTG	TGC	CCC	GCG	GGG	CAC	GCC	GTG
	CCA	GGC	GAC	AAC	ACG	GGG	CGC	CCC	GTG	CGG	CAC
3532	Gly	Ile	Phe	Arg	Ala	Ala	Val	Cys	Thr	Arg	Gly
	GGC	ATA	TTT	AGG	GCC	GCG	GTG	TGC	ACC	CGT	GGA
	CCG	TAT	AAA	TCC	CGG	CGC	CAC	ACG	TGG	GCA	CCT
3565	Val	Ala	Lys	Ala	Val	Asp	Phe	Ile	Pro	Val	Glu
	GTG	GCT	AAG	GCG	GTG	GAC	TTT	ATC	CCT	GTG	GAG
	CAC	CGA	TTC	CGC	CAC	CTG	AAA	TAG	GGA	CAC	CTC
3598	Asn	Leu	Glu	Thr	Thr	Met	Arg	Ser	Pro	Val	Phe
	AAC	CTA	GAG	ACA	ACC	ATG	AGG	TCC	CCG	GTG	TTC
	TTG	GAT	CTC	TGT	TGG	TAC	TCC	AGG	GGC	CAC	AAG
3631	Thr	Asp	Asn	Ser	Ser	Pro	Pro	Val	Val	Pro	Gln
	ACG	GAT	AAC	TCC	TCT	CCA	CCA	GTA	GTG	CCC	CAG
	TGC	CTA	TTG	AGG	AGA	GGT	GGT	CAT	CAC	GGG	GTC

FIG. 12-13

3664	Ser	Phe	Gln	Val	Ala	His	Leu	His	Ala	Pro	Thr
	AGC	TTC	CAG	GTG	GCT	CAC	CTC	CAT	GCT	CCC	ACA
	TCG	AAG	GTC	CAC	CGA	GTG	GAG	GTA	CGA	GGG	TGT
3697	Gly	Ser	Gly	Lys	Ser	Thr	Lys	Val	Pro	Ala	Ala
	GGC	AGC	GGC	AAA	AGC	ACC	AAG	GTC	CCG	GCT	GCA
	CCG	TCG	CCG	TTT	TCG	TGG	TTC	CAG	GGC	CGA	CGT
3730	Tyr	Ala	Ala	Gln	Gly	Tyr	Lys	Val	Leu	Val	Leu
	TAT	GCA	GCT	CAG	GGC	TAT	AAG	GTG	CTA	GTA	CTC
	ATA	CGT	CGA	GTC	CCG	ATA	TTC	CAC	GAT	CAT	GAG
3763	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	Phe	Gly
	AAC	CCC	TCT	GTT	GCT	GCA	ACA	CTG	GGC	TTT	GGT
	TTG	GGG	AGA	CAA	CGA	CGT	TGT	GAC	CCG	AAA	CCA
3796	Ala	Tyr	Met	Ser	Lys	Ala	His	Gly	Ile	Asp	Pro
	GCT	TAC	ATG	TCC	AAG	GCT	CAT	GGG	ATC	GAT	CCT
	CGA	ATG	TAC	AGG	TTC	CGA	GTA	CCC	TAG	CTA	GGA
3829	Asn	Ile	Arg	Thr	Gly	Val	Arg	Thr	Ile	Thr	Thr
	AAC	ATC	AGG	ACC	GGG	GTG	AGA	ACA	ATT	ACC	ACT
	TTG	TAG	TCC	TGG	CCC	CAC	TCT	TGT	TAA	TGG	TGA
3862	Gly	Ser	Pro	Ile	Thr	Tyr	Ser	Thr	Tyr	Gly	Lys
	GGC	AGC	CCC	ATC	ACG	TAC	TCC	ACC	TAC	GGC	AAG
	CCG	TCG	GGG	TAG	TGC	ATG	AGG	TGG	ATG	CCG	TTC
3895	Phe	Leu	Ala	Asp	Gly	Gly	Cys	Ser	Gly	Gly	Ala
	TTC	CTT	GCC	GAC	GGC	GGG	TGC	TCG	GGG	GGC	GCT
	AAG	GAA	CGG	CTG	CCG	CCC	ACG	AGC	CCC	CCG	CGA
3928	Tyr	Asp	Ile	Ile	Ile	Cys	Asp	Glu	Cys	His	Ser
	TAT	GAC	ATA	ATA	ATT	TGT	GAC	GAG	TGC	CAC	TCC
	ATA	CTG	TAT	TAT	TAA	ACA	CTG	CTC	ACG	GTG	AGG

FIG. 12-14

3961	Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr
	ACG GAT GCC ACA TCC ATC TTG GGC ATC GGC ACT
	TGC CTA CGG TGT AGG TAG AAC CCG TAG CCG TGA
3994	Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg
	GTC CTT GAC CAA GCA GAG ACT GCG GGG GCG AGA
	CAG GAA CTG GTT CGT CTC TGA CGC CCC CGC TCT
4027	Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly
	CTG GTT GTG CTC GCC ACC GCC ACC CCT CCG GGC
	GAC CAA CAC GAG CGG TGG CGG TGG GGA GGC CCG
4060	Ser Val Thr Val Pro His Pro Asn Ile Glu Glu
	TCC GTC ACT GTG CCC CAT CCC AAC ATC GAG GAG
	AGG CAG TGA CAC GGG GTA GGG TTG TAG CTC CTC
4093	Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe
	GTT GCT CTG TCC ACC ACC GGA GAG ATC CCT TTT
	CAA CGA GAC AGG TGG TGG CCT CTC TAG GGA AAA
4126	Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys
	TAC GGC AAG GCT ATC CCC CTC GAA GTA ATC AAG
	ATG CCG TTC CGA TAG GGG GAG CTT CAT TAG TTC
4159	Gly Gly Arg His Leu Ile Phe Cys His Ser Lys
	GGG GGG AGA CAT CTC ATC TTC TGT CAT TCA AAG
	CCC CCC TCT GTA GAG TAG AAG ACA GTA AGT TTC
4192	Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val
	AAG AAG TGC GAC GAA CTC GCC GCA AAG CTG GTC
	TTC TTC ACG CTG CTT GAG CGG CGT TTC GAC CAG
4225	Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg
	GCA TTG GGC ATC AAT GCC GTG GCC TAC TAC CGC
	CGT AAC CCG TAG TTA CGG CAC CGG ATG ATG GCG

FIG. 12-15

4258	Gly	Leu	Asp	Val	Ser	Val	Ile	Pro	Thr	Ser	Gly
	GGT	CTT	GAC	GTG	TCC	GTC	ATC	CCG	ACC	AGC	GGC
	CCA	GAA	CTG	CAC	AGG	CAG	TAG	GGC	TGG	TCG	CCG
4291	Asp	Val	Val	Val	Val	Ala	Thr	Asp	Ala	Leu	Met
	GAT	GTT	GTC	GTC	GTG	GCA	ACC	GAT	GCC	CTC	ATG
	CTA	CAA	CAG	CAG	CAC	CGT	TGG	CTA	CGG	GAG	TAC
4324	Thr	Gly	Tyr	Thr	Gly	Asp	Phe	Asp	Ser	Val	Ile
	ACC	GGC	TAT	ACC	GGC	GAC	TTC	GAC	TCG	GTG	ATA
	TGG	CCG	ATA	TGG	CCG	CTG	AAG	CTG	AGC	CAC	TAT
4357	Asp	Cys	Asn	Thr	Cys	Val	Thr	Gln	Thr	Val	Asp
	GAC	TGC	AAT	ACG	TGT	GTC	ACC	CAG	ACA	GTC	GAT
	CTG	ACG	TTA	TGC	ACA	CAG	TGG	GTC	TGT	CAG	CTA
4390	Phe	Ser	Leu	Asp	Pro	Thr	Phe	Thr	Ile	Glu	Thr
	TTC	AGC	CTT	GAC	CCT	ACC	TTC	ACC	ATT	GAG	ACA
	AAG	TCG	GAA	CTG	GGA	TGG	AAG	TGG	TAA	CTC	TGT
4423	Ile	Thr	Leu	Pro	Gln	Asp	Ala	Val	Ser	Arg	Thr
	ATC	ACG	CTC	CCC	CAG	GAT	GCT	GTC	TCC	CGC	ACT
	TAG	TGC	GAG	GGG	GTC	CTA	CGA	CAG	AGG	GCG	TGA
4456	Gln	Arg	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Lys	Pro
	CAA	CGT	CGG	GGC	AGG	ACT	GGC	AGG	GGG	AAG	CCA
	GTT	GCA	GCC	CCG	TCC	TGA	CCG	TCC	CCC	TTC	GGT
4489	Gly	Ile	Tyr	Arg	Phe	Val	Ala	Pro	Gly	Glu	Arg
	GGC	ATC	TAC	AGA	TTT	GTG	GCA	CCG	GGG	GAG	CGC
	CCG	TAG	ATG	TCT	AAA	CAC	CGT	GGC	CCC	CTC	GCG
4522	Pro	Ser	Gly	Met	Phe	Asp	Ser	Ser	Val	Leu	Cys
	CCC	TCC	GGC	ATG	TTC	GAC	TCG	TCC	GTC	CTC	TGT
	GGG	AGG	CCG	TAC	AAG	CTG	AGG	AGG	CAG	GAG	ACA

FIG. 12-16

4555	Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu
	GAG TGC TAT GAC GCA GGC TGT GCT TGG TAT GAG
	CTC ACG ATA CTG CGT CCG ACA CGA ACC ATA CTC
4588	Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg
	CTC ACG CCC GCC GAG ACT ACA GTT AGG CTA CGA
	GAG TGC GGG CGG CTC TGA TGT CAA TCC GAT GCT
4621	Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys
	GCG TAC ATG AAC ACC CCG GGG CTT CCC GTG TGC
	CGC ATG TAC TTG TGG GGC CCC GAA GGG CAC ACG
4654	Gln Asp His Leu Glu Phe Trp Glu Gly Val Phe
	CAG GAC CAT CTT GAA TTT TGG GAG GGC GTC TTT
	GTC CTG GTA GAA CTT AAA ACC CTC CCG CAG AAA
4687	Thr Gly Leu Thr His Ile Asp Ala His Phe Leu
	ACA GGC CTC ACT CAT ATA GAT GCC CAC TTT CTA
	TGT CCG GAG TGA GTA TAT CTA CGG GTG AAA GAT
4720	Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro
	TCC CAG ACA AAG CAG AGT GGG GAG AAC CTT CCT
	AGG GTC TGT TTC GTC TCA CCC CTC TTG GAA GGA
4753	Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala
	TAC CTG GTA GCG TAC CAA GCC ACC GTG TGC GCT
	ATG GAC CAT CGC ATG GTT CGG TGG CAC ACG CGA
4786	Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln
	AGG GCT CAA GCC CCT CCC CCA TCG TGG GAC CAG
	TCC CGA GTT CGG GGA GGG GGT AGC ACC CTG GTC
4819	Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
	ATG TGG AAG TGT TTG ATT CGC CTC AAG CCC ACC
	TAC ACC TTC ACA AAC TAA GCG GAG TTC GGG TGG

FIG. 12-17

4852	Leu	His	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu
	CTC	CAT	GGG	CCA	ACA	CCC	CTG	CTA	TAC	AGA	CTG
	GAG	GTA	CCC	GGT	TGT	GGG	GAC	GAT	ATG	TCT	GAC
4885	Gly	Ala	Val	Gln	Asn	Glu	Ile	Thr	Leu	Thr	His
	GGC	GCT	GTT	CAG	AAT	GAA	ATC	ACC	CTG	ACG	CAC
	CCG	CGA	CAA	GTC	TTA	CTT	TAG	TGG	GAC	TGC	GTG
4918	Pro	Val	Thr	Lys	Tyr	Ile	Met	Thr	Cys	Met	Ser
	CCA	GTC	ACC	AAA	TAC	ATC	ATG	ACA	TGC	ATG	TCG
	GGT	CAG	TGG	TTT	ATG	TAG	TAC	TGT	ACG	TAC	AGC
4951	Ala	Asp	Leu	Glu	Val	Val	Thr	Ser	Thr	Trp	Val
	GCC	GAC	CTG	GAG	GTC	GTC	ACG	AGC	ACC	TGG	GTG
	CGG	CTG	GAC	CTC	CAG	CAG	TGC	TCG	TGG	ACC	CAC
4984	Leu	Val	Gly	Gly	Val	Leu	Ala	Ala	Leu	Ala	Ala
	CTC	GTT	GGC	GGC	GTC	CTG	GCT	GCT	TTG	GCC	GCG
	GAG	CAA	CCG	CCG	CAG	GAC	CGA	CGA	AAC	CGG	CGC
5017	Tyr	Cys	Leu	Ser	Thr	Gly	Cys	Val	Val	Ile	Val
	TAT	TGC	CTG	TCA	ACA	GGC	TGC	GTG	GTC	ATA	GTG
	ATA	ACG	GAC	AGT	TGT	CCG	ACG	CAC	CAG	TAT	CAC
5050	Gly	Arg	Val	Val	Leu	Ser	Gly	Lys	Pro	Ala	Ile
	GGC	AGG	GTC	GTC	TTG	TCC	GGG	AAG	CCG	GCA	ATC
	CCG	TCC	CAG	CAG	AAC	AGG	CCC	TTC	GGC	CGT	TAG
5083	Ile	Pro	Asp	Arg	Glu	Val	Leu	Tyr	Arg	Glu	Phe
	ATA	CCT	GAC	AGG	GAA	GTC	CTC	TAC	CGA	GAG	TTC
	TAT	GGA	CTG	TCC	CTT	CAG	GAG	ATG	GCT	CTC	AAG
5116	Asp	Glu	Met	Glu	Glu	Cys	Ser	Gln	His	Leu	Pro
	GAT	GAG	ATG	GAA	GAG	TGC	TCT	CAG	CAC	TTA	CCG
	CTA	CTC	TAC	CTT	CTC	ACG	AGA	GTC	GTG	AAT	GGC

FIG. 12-18

5149	Tyr	Ile	Glu	Gln	Gly	Met	Met	Leu	Ala	Glu	Gln
	TAC	ATC	GAG	CAA	GGG	ATG	ATG	CTC	GCC	GAG	CAG
	ATG	TAG	CTC	GTT	CCC	TAC	TAC	GAG	CGG	CTC	GTC
5182	Phe	Lys	Gln	Lys	Ala	Leu	Gly	Leu	Leu	Gln	Thr
	TTC	AAG	CAG	AAG	GCC	CTC	GGC	CTC	CTG	CAG	ACC
	AAG	TTC	GTC	TTC	CGG	GAG	CCG	GAG	GAC	GTC	TGG
5215	Ala	Ser	Arg	Gln	Ala	Glu	Val	Ile	Ala	Pro	Ala
	GCG	TCC	CGT	CAG	GCA	GAG	GTT	ATC	GCC	CCT	GCT
	CGC	AGG	GCA	GTC	CGT	CTC	CAA	TAG	CGG	GGA	CGA
5248	Val	Gln	Thr	Asn	Trp	Gln	Lys	Leu	Glu	Thr	Phe
	GTC	CAG	ACC	AAC	TGG	CAA	AAA	CTC	GAG	ACC	TTC
	CAG	GTC	TGG	TTG	ACC	GTT	TTT	GAG	CTC	TGG	AAG
5281	Trp	Ala	Lys	His	Met	Trp	Asn	Phe	Ile	Ser	Gly
	TGG	GCG	AAG	CAT	ATG	TGG	AAC	TTC	ATC	AGT	GGG
	ACC	CGC	TTC	GTA	TAC	ACC	TTG	AAG	TAG	TCA	CCC
5314	Ile	Gln	Tyr	Leu	Ala	Gly	Leu	Ser	Thr	Leu	Pro
	ATA	CAA	TAC	TTG	GCG	GGC	TTG	TCA	ACG	CTG	CCT
	TAT	GTT	ATG	AAC	CGC	CCG	AAC	AGT	TGC	GAC	GGA
5347	Gly	Asn	Pro	Ala	Ile	Ala	Ser	Leu	Met	Ala	Phe
	GGT	AAC	CCC	GCC	ATT	GCT	TCA	TTG	ATG	GCT	TTT
	CCA	TTG	GGG	CGG	TAA	CGA	AGT	AAC	TAC	CGA	AAA
5380	Thr	Ala	Ala	Val	Thr	Ser	Pro	Leu	Thr	Thr	Ser
	ACA	GCT	GCT	GTC	ACC	AGC	CCA	CTA	ACC	ACT	AGC
	TGT	CGA	CGA	CAG	TGG	TCG	GGT	GAT	TGG	TGA	TCG
5413	Gln	Thr	Leu	Leu	Phe	Asn	Ile	Leu	Gly	Gly	Trp
	CAA	ACC	CTC	CTC	TTC	AAC	ATA	TTG	GGG	GGG	TGG
	GTT	TGG	GAG	GAG	AAG	TTG	TAT	AAC	CCC	CCC	ACC

FIG. 12-19

5446	Val	Ala	Ala	Gln	Leu	Ala	Ala	Pro	Gly	Ala	Ala
	GTG	GCT	GCC	CAG	CTC	GCC	GCC	CCC	GGT	GCC	GCT
	CAC	CGA	CGG	GTC	GAG	CGG	CGG	GGG	CCA	CGG	CGA
5479	Thr	Ala	Phe	Val	Gly	Ala	Gly	Leu	Ala	Gly	Ala
	ACT	GCC	TTT	GTG	GGC	GCT	GGC	TTA	GCT	GGC	GCC
	TGA	CGG	AAA	CAC	CCG	CGA	CCG	AAT	CGA	CCG	CGG
5512	Ala	Ile	Gly	Ser	Val	Gly	Leu	Gly	Lys	Val	Leu
	GCC	ATC	GGC	AGT	GTT	GGA	CTG	GGG	AAG	GTC	CTC
	CGG	TAG	CCG	TCA	CAA	CCT	GAC	CCC	TTC	CAG	GAG
5545	Ile	Asp	Ile	Leu	Ala	Gly	Tyr	Gly	Ala	Gly	Val
	ATA	GAC	ATC	CTT	GCA	GGG	TAT	GGC	GCG	GGC	GTG
	TAT	CTG	TAG	GAA	CGT	CCC	ATA	CCG	CGC	CCG	CAC
5578	Ala	Gly	Ala	Leu	Val	Ala	Phe	Lys	Ile	Met	Ser
	GCG	GGA	GCT	CTT	GTG	GCA	TTC	AAG	ATC	ATG	AGC
	CGC	CCT	CGA	GAA	CAC	CGT	AAG	TTC	TAG	TAC	TCG
5611	Gly	Glu	Val	Pro	Ser	Thr	Glu	Asp	Leu	Val	Asn
	GGT	GAG	GTC	CCC	TCC	ACG	GAG	GAC	CTG	GTC	AAT
	CCA	CTC	CAG	GGG	AGG	TGC	CTC	CTG	GAC	CAG	TTA
5644	Leu	Leu	Pro	Ala	Ile	Leu	Ser	Pro	Gly	Ala	Leu
	CTA	CTG	CCC	GCC	ATC	CTC	TCG	CCC	GGA	GCC	CTC
	GAT	GAC	GGG	CGG	TAG	GAG	AGC	GGG	CCT	CGG	GAG
5677	Val	Val	Gly	Val	Val	Cys	Ala	Ala	Ile	Leu	Arg
	GTA	GTC	GGC	GTG	GTC	TGT	GCA	GCA	ATA	CTG	CGC
	CAT	CAG	CCG	CAC	CAG	ACA	CGT	CGT	TAT	GAC	GCG
5710	Arg	His	Val	Gly	Pro	Gly	Glu	Gly	Ala	Val	Gln
	CGG	CAC	GTT	GGC	CCG	GGC	GAG	GGG	GCA	GTG	CAG
	GCC	GTG	CAA	CCG	GGC	CCG	CTC	CCC	CGT	CAC	GTC

FIG. 12-20

5743	Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg
	TGG ATG AAC CGG CTG ATA GCC TTC GCC TCC CGG
	ACC TAC TTG GCC GAC TAT CGG AAG CGG AGG GCC
5776	Gly Asn His Val Ser Pro Thr His Tyr Val Pro
	GGG AAC CAT GTT TCC CCC ACG CAC TAC GTG CCG
	CCC TTG GTA CAA AGG GGG TGC GTG ATG CAC GGC
5809	Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile
	GAG AGC GAT GCA GCT GCC CGC GTC ACT GCC ATA
	CTC TCG CTA CGT CGA CGG GCG CAG TGA CGG TAT
5842	Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg
	CTC AGC AGC CTC ACT GTA ACC CAG CTC CTG AGG
	GAG TCG TCG GAG TGA CAT TGG GTC GAG GAC TCC
5875	Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr
	CGA CTG CAC CAG TGG ATA AGC TCG GAG TGT ACC
	GCT GAC GTG GTC ACC TAT TCG AGC CTC ACA TGG
5908	Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile
	ACT CCA TGC TCC GGT TCC TGG CTA AGG GAC ATC
	TGA GGT ACG AGG CCA AGG ACC GAT TCC CTG TAG
5941	Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe
	TGG GAC TGG ATA TGC GAG GTG TTG AGC GAC TTT
	ACC CTG ACC TAT ACG CTC CAC AAC TCG CTG AAA
5974	Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln
	AAG ACC TGG CTA AAA GCT AAG CTC ATG CCA CAG
	TTC TGG ACC GAT TTT CGA TTC GAG TAC GGT GTC
6007	Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg
	CTG CCT GGG ATC CCC TTT GTG TCC TGC CAG CGC
	GAC GGA CCC TAG GGG AAA CAC AGG ACG GTC GCG

FIG. 12-21

6040	Gly Tyr Lys Gly Val Trp Arg Val Asp Gly Ile
	GGG TAT AAG GGG GTC TGG CGA GTG GAC GGC ATC
	CCC ATA TTC CCC CAG ACC GCT CAC CTG CCG TAG
6073	Met His Thr Arg Cys His Cys Gly Ala Glu Ile
	ATG CAC ACT CGC TGC CAC TGT GGA GCT GAG ATC
	TAC GTG TGA GCG ACG GTG ACA CCT CGA CTC TAG
6106	Thr Gly His Val Lys Asn Gly Thr Met Arg Ile
	ACT GGA CAT GTC AAA AAC GGG ACG ATG AGG ATC
	TGA CCT GTA CAG TTT TTG CCC TGC TAC TCC TAG
6139	Val Gly Pro Arg Thr Cys Arg Asn Met Trp Ser
	GTC GGT CCT AGG ACC TGC AGG AAC ATG TGG AGT
	CAG CCA GGA TCC TGG ACG TCC TTG TAC ACC TCA
6172	Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly
	GGG ACC TTC CCC ATT AAT GCC TAC ACC ACG GGC
	CCC TGG AAG GGG TAA TTA CGG ATG TGG TGC CCG
6205	Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr
	CCC TGT ACC CCC CTT CCT GCG CCG AAC TAC ACG
	GGG ACA TGG GGG GAA GGA CGC GGC TTG ATG TGC
6238	Phe Ala Leu Trp Arg Val Ser Ala Glu Glu Tyr
	TTC GCG CTA TGG AGG GTG TCT GCA GAG GAA TAT
	AAG CGC GAT ACC TCC CAC AGA CGT CTC CTT ATA
6271	Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr
	GTG GAG ATA AGG CAG GTG GGG GAC TTC CAC TAC
	CAC CTC TAT TCC GTC CAC CCC CTG AAG GTG ATG
6304	Val Thr Gly Met Thr Thr Asp Asn Leu Lys Cys
	GTG ACG GGT ATG ACT ACT GAC AAT CTC AAA TGC
	CAC TGC CCA TAC TGA TGA CTG TTA GAG TTT ACG

FIG. 12-22

6337	Pro	Cys	Gln	Val	Pro	Ser	Pro	Glu	Phe	Phe	Thr
	CCG	TGC	CAG	GTC	CCA	TCG	CCC	GAA	TTT	TTC	ACA
	GGC	ACG	GTC	CAG	GGT	AGC	GGG	CTT	AAA	AAG	TGT
6370	Glu	Leu	Asp	Gly	Val	Arg	Leu	His	Arg	Phe	Ala
	GAA	TTG	GAC	GGG	GTG	CGC	CTA	CAT	AGG	TTT	GCG
	CTT	AAC	CTG	CCC	CAC	GCG	GAT	GTA	TCC	AAA	CGC
6403	Pro	Pro	Cys	Lys	Pro	Leu	Leu	Arg	Glu	Glu	Val
	CCC	CCC	TGC	AAG	CCC	TTG	CTG	CGG	GAG	GAG	GTA
	GGG	GGG	ACG	TTC	GGG	AAC	GAC	GCC	CTC	CTC	CAT
6436	Ser	Phe	Arg	Val	Gly	Leu	His	Glu	Tyr	Pro	Val
	TCA	TTC	AGA	GTA	GGA	CTC	CAC	GAA	TAC	CCG	GTA
	AGT	AAG	TCT	CAT	CCT	GAG	GTG	CTT	ATG	GGC	CAT
6469	Gly	Ser	Gln	Leu	Pro	Cys	Glu	Pro	Glu	Pro	Asp
	GGG	TCG	CAA	TTA	CCT	TGC	GAG	CCC	GAA	CCG	GAC
	CCC	AGC	GTT	AAT	GGA	ACG	CTC	GGG	CTT	GGC	CTG
6502	Val	Ala	Val	Leu	Thr	Ser	Met	Leu	Thr	Asp	Pro
	GTG	GCC	GTG	TTG	ACG	TCC	ATG	CTC	ACT	GAT	CCC
	CAC	CGG	CAC	AAC	TGC	AGG	TAC	GAG	TGA	CTA	GGG
6535	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Gly	Arg	Arg
	TCC	CAT	ATA	ACA	GCA	GAG	GCG	GCC	GGG	CGA	AGG
	AGG	GTA	TAT	TGT	CGT	CTC	CGC	CGG	CCC	GCT	TCC
6568	Leu	Ala	Arg	Gly	Ser	Pro	Pro	Ser	Val	Ala	Ser
	TTG	GCG	AGG	GGA	TCA	CCC	CCC	TCT	GTG	GCC	AGC
	AAC	CGC	TCC	CCT	AGT	GGG	GGG	AGA	CAC	CGG	TCG
6601	Ser	Ser	Ala	Ser	Gln	Leu	Ser	Ala	Pro	Ser	Leu
	TCC	TCG	GCT	AGC	CAG	CTA	TCC	GCT	CCA	TCT	CTC
	AGG	AGC	CGA	TCG	GTC	GAT	AGG	CGA	GGT	AGA	GAG

FIG. 12-23

6634	Lys	Ala	Thr	Cys	Thr	Ala	Asn	His	Asp	Ser	Pro
	AAG	GCA	ACT	TGC	ACC	GCT	AAC	CAT	GAC	TCC	CCT
	TTC	CGT	TGA	ACG	TGG	CGA	TTG	GTA	CTG	AGG	GGA
6667	Asp	Ala	Glu	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp
	GAT	GCT	GAG	CTC	ATA	GAG	GCC	AAC	CTC	CTA	TGG
	CTA	CGA	CTC	GAG	TAT	CTC	CGG	TTG	GAG	GAT	ACC
6700	Arg	Gln	Glu	Met	Gly	Gly	Asn	Ile	Thr	Arg	Val
	AGG	CAG	GAG	ATG	GGC	GGC	AAC	ATC	ACC	AGG	GTT
	TCC	GTC	CTC	TAC	CCG	CCG	TTG	TAG	TGG	TCC	CAA
6733	Glu	Ser	Glu	Asn	Lys	Val	Val	Ile	Leu	Asp	Ser
	GAG	TCA	GAA	AAC	AAA	GTG	GTG	ATT	CTG	GAC	TCC
	CTC	AGT	CTT	TTG	TTT	CAC	CAC	TAA	GAC	CTG	AGG
6766	Phe	Asp	Pro	Leu	Val	Ala	Glu	Glu	Asp	Glu	Arg
	TTC	GAT	CCG	CTT	GTG	GCG	GAG	GAG	GAC	GAG	CGG
	AAG	CTA	GGC	GAA	CAC	CGC	CTC	CTC	CTG	CTC	GCC
6799	Glu	Ile	Ser	Val	Pro	Ala	Glu	Ile	Leu	Arg	Lys
	GAG	ATC	TCC	GTA	CCC	GCA	GAA	ATC	CTG	CGG	AAG
	CTC	TAG	AGG	CAT	GGG	CGT	CTT	TAG	GAC	GCC	TTC
6832	Ser	Arg	Arg	Phe	Ala	Gln	Ala	Leu	Pro	Val	Trp
	TCT	CGG	AGA	TTC	GCC	CAG	GCC	CTG	CCC	GTT	TGG
	AGA	GCC	TCT	AAG	CGG	GTC	CGG	GAC	GGG	CAA	ACC
6865	Ala	Arg	Pro	Asp	Tyr	Asn	Pro	Pro	Leu	Val	Glu
	GCG	CGG	CCG	GAC	TAT	AAC	CCC	CCG	CTA	GTG	GAG
	CGC	GCC	GGC	CTG	ATA	TTG	GGG	GGC	GAT	CAC	CTC
6898	Thr	Trp	Lys	Lys	Pro	Asp	Tyr	Glu	Pro	Pro	Val
	ACG	TGG	AAA	AAG	CCC	GAC	TAC	GAA	CCA	CCT	GTG
	TGC	ACC	TTT	TTC	GGG	CTG	ATG	CTT	GGT	GGA	CAC

FIG. 12-24

6931	Val	His	Gly	Cys	Pro	Leu	Pro	Pro	Pro	Lys	Ser
	GTC	CAT	GGC	TGT	CCG	CTT	CCA	CCT	CCA	AAG	TCC
	CAG	GTA	CCG	ACA	GGC	GAA	GGT	GGA	GGT	TTC	AGG
6964	Pro	Pro	Val	Pro	Pro	Pro	Arg	Lys	Lys	Arg	Thr
	CCT	CCT	GTG	CCT	CCG	CCT	CGG	AAG	AAG	CGG	ACG
	GGA	GGA	CAC	GGA	GGC	GGA	GCC	TTC	TTC	GCC	TGC
6997	Val	Val	Leu	Thr	Glu	Ser	Thr	Leu	Ser	Thr	Ala
	GTG	GTC	CTC	ACT	GAA	TCA	ACC	CTA	TCT	ACT	GCC
	CAC	CAG	GAG	TGA	CTT	AGT	TGG	GAT	AGA	TGA	CGG
7030	Leu	Ala	Glu	Leu	Ala	Thr	Arg	Ser	Phe	Gly	Ser
	TTG	GCC	GAG	CTC	GCC	ACC	AGA	AGC	TTT	GGC	AGC
	AAC	CGG	CTC	GAG	CGG	TGG	TCT	TCG	AAA	CCG	TCG
7063	Ser	Ser	Thr	Ser	Gly	Ile	Thr	Gly	Asp	Asn	Thr
	TCC	TCA	ACT	TCC	GGC	ATT	ACG	GGC	GAC	AAT	ACG
	AGG	AGT	TGA	AGG	CCG	TAA	TGC	CCG	CTG	TTA	TGC
7096	Thr	Thr	Ser	Ser	Glu	Pro	Ala	Pro	Ser	Gly	Cys
	ACA	ACA	TCC	TCT	GAG	CCC	GCC	CCT	TCT	GGC	TGC
	TGT	TGT	AGG	AGA	CTC	GGG	CGG	GGA	AGA	CCG	ACG
7129	Pro	Pro	Asp	Ser	Asp	Ala	Glu	Ser	Tyr	Ser	Ser
	CCC	CCC	GAC	TCC	GAC	GCT	GAG	TCC	TAT	TCC	TCC
	GGG	GGG	CTG	AGG	CTG	CGA	CTC	AGG	ATA	AGG	AGG
7162	Met	Pro	Pro	Leu	Glu	Gly	Glu	Pro	Gly	Asp	Pro
	ATG	CCC	CCC	CTG	GAG	GGG	GAG	CCT	GGG	GAT	CCG
	TAC	GGG	GGG	GAC	CTC	CCC	CTC	GGA	CCC	CTA	GGC
7195	Asp	Leu	Ser	Asp	Gly	Ser	Trp	Ser	Thr	Val	Ser
	GAT	CTT	AGC	GAC	GGG	TCA	TGG	TCA	ACG	GTC	AGT
	CTA	GAA	TCG	CTG	CCC	AGT	ACC	AGT	TGC	CAG	TCA

FIG. 12-25

7228	Ser	Glu	Ala	Asn	Ala	Glu	Asp	Val	Val	Cys	Cys
	AGT	GAG	GCC	AAC	GCG	GAG	GAT	GTC	GTG	TGC	TGC
	TCA	CTC	CGG	TTG	CGC	CTC	CTA	CAG	CAC	ACG	ACG
7261	Ser	Met	Ser	Tyr	Ser	Trp	Thr	Gly	Ala	Leu	Val
	TCA	ATG	TCT	TAC	TCT	TGG	ACA	GGC	GCA	CTC	GTC
	AGT	TAC	AGA	ATG	AGA	ACC	TGT	CCG	CGT	GAG	CAG
7294	Thr	Pro	Cys	Ala	Ala	Glu	Glu	Gln	Lys	Leu	Pro
	ACC	CCG	TGC	GCC	GCG	GAA	GAA	CAG	AAA	CTG	CCC
	TGG	GGC	ACG	CGG	CGC	CTT	CTT	GTC	TTT	GAC	GGG
7327	Ile	Asn	Ala	Leu	Ser	Asn	Ser	Leu	Leu	Arg	His
	ATC	AAT	GCA	CTA	AGC	AAC	TCG	TTG	CTA	CGT	CAC
	TAG	TTA	CGT	GAT	TCG	TTG	AGC	AAC	GAT	GCA	GTG
7360	His	Asn	Leu	Val	Tyr	Ser	Thr	Thr	Ser	Arg	Ser
	CAC	AAT	TTG	GTG	TAT	TCC	ACC	ACC	TCA	CGC	AGT
	GTG	TTA	AAC	CAC	ATA	AGG	TGG	TGG	AGT	GCG	TCA
7393	Ala	Cys	Gln	Arg	Gln	Lys	Lys	Val	Thr	Phe	Asp
	GCT	TGC	CAA	AGG	CAG	AAG	AAA	GTC	ACA	TTT	GAC
	CGA	ACG	GTT	TCC	GTC	TTC	TTT	CAG	TGT	AAA	CTG
7426	Arg	Leu	Gln	Val	Leu	Asp	Ser	His	Tyr	Gln	Asp
	AGA	CTG	CAA	GTT	CTG	GAC	AGC	CAT	TAC	CAG	GAC
	TCT	GAC	GTT	CAA	GAC	CTG	TCG	GTA	ATG	GTC	CTG
7459	Val	Leu	Lys	Glu	Val	Lys	Ala	Ala	Ala	Ser	Lys
	GTA	CTC	AAG	GAG	GTT	AAA	GCA	GCG	GCG	TCA	AAA
	CAT	GAG	TTC	CTC	CAA	TTT	CGT	CGC	CGC	AGT	TTT
7492	Val	Lys	Ala	Asn	Leu	Leu	Ser	Val	Glu	Glu	Ala
	GTG	AAG	GCT	AAC	TTG	CTA	TCC	GTA	GAG	GAA	GCT
	CAC	TTC	CGA	TTG	AAC	GAT	AGG	CAT	CTC	CTT	CGA

FIG. 12-26

7525	Cys	Ser	Leu	Thr	Pro	Pro	His	Ser	Ala	Lys	Ser
	TGC	AGC	CTG	ACG	CCC	CCA	CAC	TCA	GCC	AAA	TCC
	ACG	TCG	GAC	TGC	GGG	GGT	GTG	AGT	CGG	TTT	AGG
7558	Lys	Phe	Gly	Tyr	Gly	Ala	Lys	Asp	Val	Arg	Cys
	AAG	TTT	GGT	TAT	GGG	GCA	AAA	GAC	GTC	CGT	TGC
	TTC	AAA	CCA	ATA	CCC	CGT	TTT	CTG	CAG	GCA	ACG
7591	His	Ala	Arg	Lys	Ala	Val	Thr	His	Ile	Asn	Ser
	CAT	GCC	AGA	AAG	GCC	GTA	ACC	CAC	ATC	AAC	TCC
	GTA	CGG	TCT	TTC	CGG	CAT	TGG	GTG	TAG	TTG	AGG
7624	Val	Trp	Lys	Asp	Leu	Leu	Glu	Asp	Asn	Val	Thr
	GTG	TGG	AAA	GAC	CTT	CTG	GAA	GAC	AAT	GTA	ACA
	CAC	ACC	TTT	CTG	GAA	GAC	CTT	CTG	TTA	CAT	TGT
7657	Pro	Ile	Asp	Thr	Thr	Ile	Met	Ala	Lys	Asn	Glu
	CCA	ATA	GAC	ACT	ACC	ATC	ATG	GCT	AAG	AAC	GAG
	GGT	TAT	CTG	TGA	TGG	TAG	TAC	CGA	TTC	TTG	CTC
7690	Val	Phe	Cys	Val	Gln	Pro	Glu	Lys	Gly	Gly	Arg
	GTT	TTC	TGC	GTT	CAG	CCT	GAG	AAG	GGG	GGT	CGT
	CAA	AAG	ACG	CAA	GTC	GGA	CTC	TTC	CCC	CCA	GCA
7723	Lys	Pro	Ala	Arg	Leu	Ile	Val	Phe	Pro	Asp	Leu
	AAG	CCA	GCT	CGT	CTC	ATC	GTG	TTC	CCC	GAT	CTG
	TTC	GGT	CGA	GCA	GAG	TAG	CAC	AAG	GGG	CTA	GAC
7756	Gly	Val	Arg	Val	Cys	Glu	Lys	Met	Ala	Leu	Tyr
	GGC	GTG	CGC	GTG	TGC	GAA	AAG	ATG	GCT	TTG	TAC
	CCG	CAC	GCG	CAC	ACG	CTT	TTC	TAC	CGA	AAC	ATG
7789	Asp	Val	Val	Thr	Lys	Leu	Pro	Leu	Ala	Val	Met
	GAC	GTG	GTT	ACA	AAG	CTC	CCC	TTG	GCC	GTG	ATG
	CTG	CAC	CAA	TGT	TTC	GAG	GGG	AAC	CGG	CAC	TAC

FIG. 12-27

7822	Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly
	GGA AGC TCC TAC GGA TTC CAA TAC TCA CCA GGA
	CCT TCG AGG ATG CCT AAG GTT ATG AGT GGT CCT
7855	Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys
	CAG CGG GTT GAA TTC CTC GTG CAA GCG TGG AAG
	GTC GCC CAA CTT AAG GAG CAC GTT CGC ACC TTC
7888	Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp
	TCC AAG AAA ACC CCA ATG GGG TTC TCG TAT GAT
	AGG TTC TTT TGG GGT TAC CCC AAG AGC ATA CTA
7921	Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser
	ACC CGC TGC TTT GAC TCC ACA GTC ACT GAG AGC
	TGG GCG ACG AAA CTG AGG TGT CAG TGA CTC TCG
7954	Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys
	GAC ATC CGT ACG GAG GAG GCA ATC TAC CAA TGT
	CTG TAG GCA TGC CTC CTC CGT TAG ATG GTT ACA
7987	Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile
	TGT GAC CTC GAC CCC CAA GCC CGC GTG GCC ATC
	ACA CTG GAG CTG GGG GTT CGG GCG CAC CGG TAG
8020	Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly
	AAG TCC CTC ACC GAG AGG CTT TAT GTT GGG GGC
	TTC AGG GAG TGG CTC TCC GAA ATA CAA CCC CCG
8053	Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly
	CCT CTT ACC AAT TCA AGG GGG GAG AAC TGC GGC
	GGA GAA TGG TTA AGT TCC CCC CTC TTG ACG CCG
8086	Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr
	TAT CGC AGG TGC CGC GCG AGC GGC GTA CTG ACA
	ATA GCG TCC ACG GCG CGC TCG CCG CAT GAC TGT

FIG. 12-28

8119	Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile
	ACT AGC TGT GGT AAC ACC CTC ACT TGC TAC ATC
	TGA TCG ACA CCA TTG TGG GAG TGA ACG ATG TAG
8152	Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
	AAG GCC CGG GCA GCC TGT CGA GCC GCA GGG CTC
	TTC CGG GCC CGT CGG ACA GCT CGG CGT CCC GAG
8185	Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp
	CAG GAC TGC ACC ATG CTC GTG TGT GGC GAC GAC
	GTC CTG ACG TGG TAC GAG CAC ACA CCG CTG CTG
8218	Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln
	TTA GTC GTT ATC TGT GAA AGC GCG GGG GTC CAG
	AAT CAG CAA TAG ACA CTT TCG CGC CCC CAG GTC
8251	Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr Glu
	GAG GAC GCG GCG AGC CTG AGA GCC TTC ACG GAG
	CTC CTG CGC CGC TCG GAC TCT CGG AAG TGC CTC
8284	Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp
	GCT ATG ACC AGG TAC TCC GCC CCC CCT GGG GAC
	CGA TAC TGG TCC ATG AGG CGG GGG GGA CCC CTG
8317	Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile
	CCC CCA CAA CCA GAA TAC GAC TTG GAG CTC ATA
	GGG GGT GTT GGT CTT ATG CTG AAC CTC GAG TAT
8350	Thr Ser Cys Ser Ser Asn Val Ser Val Ala His
	ACA TCA TGC TCC TCC AAC GTG TCA GTC GCC CAC
	TGT AGT ACG AGG AGG TTG CAC AGT CAG CGG GTG
8383	Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr
	GAC GGC GCT GGA AAG AGG GTC TAC TAC CTC ACC
	CTG CCG CGA CCT TTC TCC CAG ATG ATG GAG TGG

FIG. 12-29

8416	Arg	Asp	Pro	Thr	Thr	Pro	Leu	Ala	Arg	Ala	Ala
	CGT	GAC	CCT	ACA	ACC	CCC	CTC	GCG	AGA	GCT	GCG
	GCA	CTG	GGA	TGT	TGG	GGG	GAG	CGC	TCT	CGA	CGC
8449	Trp	Glu	Thr	Ala	Arg	His	Thr	Pro	Val	Asn	Ser
	TGG	GAG	ACA	GCA	AGA	CAC	ACT	CCA	GTC	AAT	TCC
	ACC	CTC	TGT	CGT	TCT	GTG	TGA	GGT	CAG	TTA	AGG
8482	Trp	Leu	Gly	Asn	Ile	Ile	Met	Phe	Ala	Pro	Thr
	TGG	CTA	GGC	AAC	ATA	ATC	ATG	TTT	GCC	CCC	ACA
	ACC	GAT	CCG	TTG	TAT	TAG	TAC	AAA	CGG	GGG	TGT
8515	Leu	Trp	Ala	Arg	Met	Ile	Leu	Met	Thr	His	Phe
	CTG	TGG	GCG	AGG	ATG	ATA	CTG	ATG	ACC	CAT	TTC
	GAC	ACC	CGC	TCC	TAC	TAT	GAC	TAC	TGG	GTA	AAG
8548	Phe	Ser	Val	Leu	Ile	Ala	Arg	Asp	Gln	Leu	Glu
	TTT	AGC	GTC	CTT	ATA	GCC	AGG	GAC	CAG	CTT	GAA
	AAA	TCG	CAG	GAA	TAT	CGG	TCC	CTG	GTC	GAA	CTT
8581	Gln	Ala	Leu	Asp	Cys	Glu	Ile	Tyr	Gly	Ala	Cys
	CAG	GCC	CTC	GAT	TGC	GAG	ATC	TAC	GGG	GCC	TGC
	GTC	CGG	GAG	CTA	ACG	CTC	TAG	ATG	CCC	CGG	ACG
8614	Tyr	Ser	Ile	Glu	Pro	Leu	Asp	Leu	Pro	Pro	Ile
	TAC	TCC	ATA	GAA	CCA	CTT	GAT	CTA	CCT	CCA	ATC
	ATG	AGG	TAT	CTT	GGT	GAA	CTA	GAT	GGA	GGT	TAG
8647	Ile	Gln	Arg	Leu	His	Gly	Leu	Ser	Ala	Phe	Ser
	ATT	CAA	AGA	CTC	CAT	GGC	CTC	AGC	GCA	TTT	TCA
	TAA	GTT	TCT	GAG	GTA	CCG	GAG	TCG	CGT	AAA	AGT
8680	Leu	His	Ser	Tyr	Ser	Pro	Gly	Glu	Ile	Asn	Arg
	CTC	CAC	AGT	TAC	TCT	CCA	GGT	GAA	ATT	AAT	AGG
	GAG	GTG	TCA	ATG	AGA	GGT	CCA	CTT	TAA	TTA	TCC

FIG. 12-30

8713	Val	Ala	Ala	Cys	Leu	Arg	Lys	Leu	Gly	Val	Pro
	GTG	GCC	GCA	TGC	CTC	AGA	AAA	CTT	GGG	GTA	CCG
	CAC	CGG	CGT	ACG	GAG	TCT	TTT	GAA	CCC	CAT	GGC
8746	Pro	Leu	Arg	Ala	Trp	Arg	His	Arg	Ala	Arg	Ser
	CCC	TTG	CGA	GCT	TGG	AGA	CAC	CGG	GCC	CGG	AGC
	GGG	AAC	GCT	CGA	ACC	TCT	GTG	GCC	CGG	GCC	TCG
8779	Val	Arg	Ala	Arg	Leu	Leu	Ala	Arg	Gly	Gly	Arg
	GTC	CGC	GCT	AGG	CTT	CTG	GCC	AGA	GGA	GGC	AGG
	CAG	GCG	CGA	TCC	GAA	GAC	CGG	TCT	CCT	CCG	TCC
8812	Ala	Ala	Ile	Cys	Gly	Lys	Tyr	Leu	Phe	Asn	Trp
	GCT	GCC	ATA	TGT	GGC	AAG	TAC	CTC	TTC	AAC	TGG
	CGA	CGG	TAT	ACA	CCG	TTC	ATG	GAG	AAG	TTG	ACC
8845	Ala	Val	Arg	Thr	Lys	Leu	Lys				
	GCA	GTA	AGA	ACA	AAG	CTC	AAA	C			
	CGT	CAT	TCT	TGT	TTC	GAG	TTT	G			

FIG. 12-31

J1
PT

primer J159S
ACTGCCCTGAACTGCAATGA
G

			Ser	Leu	Lys	Thr	Gly	Phe	Leu	Ala	Ala
J1	1	C	TCC	CTC	AAA	ACT	GGG	TTT	CTT	GCC	GCG
PT		T	AG		C	C	C	GG	T G	A	G
					<u>Asn</u>			Trp			Gly

J1	29	Leu	Phe	Tyr	Thr	His	Lys	Phe	Asn	Ala
PT		CTG	TTC	TAC	ACA	CAC	AAG	TTC	AAC	GCG
		T		T	CAC					T T
				<u>His</u>						Ser

primer 166A for J1-1216

[illegible]

J1	83	Cys	Arg	Ser	Ile	Asp	Lys	Phe	Asp	Gln
PT		TGT	CGC	TCC	ATT	GAC	AAG	TTC	GAC	CAG
		C	A	C	C	AC	G T	T		
				Pro	Leu	<u>Thr</u>	<u>Asp</u>			

J1	110	Gly	Trp	Gly	Pro	Ile	Thr	Tyr	Ala	Gln
PT		GGG	TGG	GGT	CCC	ATC	ACC	TAT	GCT	CAA
		C		C	T		GT		C	A C
							Ser			Asn

J1	137	Pro	Asp	Asn	Ser	Asp	Gln	Arg	Pro	Tyr
PT		CCT	GAC	AAC	TCG	GAC	CAG	AGG	CCG	TAT
		GGA	AG	GG	C C			C C	C	C
		Gly	<u>Ser</u>	<u>Gly</u>	Pro					

J1	164	Cys	Trp	His	Tyr	Ala	Pro	Arg	Gln	Cys
PT		TGC	TGG	CAC	TAC	GCA	CCT	CGA	CAG	TGT
						C C	A	AA	CT	C
						Pro		Lys	Pro	

FIG. 13-1

J1	191	Gly	Ile	Val	Pro	Ala	Ser	Gln	Val	Cys
PT		GGT	ATC	GTA	CCC	GCG	TCG	CAG	GTG	TGC
			T	G			AA	AGT		T
							<u>Lys</u>	<u>Ser</u>		
J1	218	Gly	Pro	Val	Tyr	Cys	Phe	Thr	Pro	Ser
PT		GGT	CCA	GTG	TAT	TGC	TTC	ACC	CCA	AGC
			G	A				T	C	
J1	245	Pro	Val	Val	Val	Gly	Thr	Thr	Asp	Arg
PT		CCT	GTT	GTA	GTG	GGG	ACG	ACC	GAT	CGT
		C	G	G		A			C	A G
J1	272	Phe	Gly	Ala	Pro	Thr	Tyr	Asn	Trp	Gly
PT		TTC	GGC	GCC	CCT	ACG	TAT	AAC	TGG	GGG
		CG		G	C	C	C	G		T
		<u>Ser</u>						<u>Ser</u>		
J1	299	Asp	Asn	Glu	Thr	Asp	Val	Leu	Leu	Leu
PT		GAC	AAT	GAG	ACG	GAC	GTG	CTG	CTC	CTA
		A		T			C	T	C	T
		Glu		Asp				Phe	Val	
J1	326	Asn	Asn	Thr	Arg	Pro	Pro	His	Gly	Asn
PT		AAC	AAC	ACG	CGG	CCC	CCG	CAC	GGC	AAC
			T	C	A	A		TG		T
								<u>Leu</u>		
J1	353	Trp	Phe	Gly	Cys	Thr				
PT		TGG	TTC	GGC	TGT	ACA				
				T						
								<u>CTGGATGAACTCAACTGGATT</u>		
								primer 199A		

Nucleotide Match: 259/367 (70.6%)
 Amino Acid Match (stringent): 93/122 (76.2%)
 (relaxed): 111/122 (91.0%)

FIG. 13-2

Prototype HCV (PT) sequences different from
Japanese HCV (J1) are shown.
Relaxed amino acid match: Gly=Ala=Pro=Ser=Thr,
Asp=Glu, Asn=Gln,
 Aug=Lys=His, Leu=Ile=Val=Met, Phe=Trp=Tyr.
Underline, different amino acid in relaxed
matching.

FIG. 13-3

Core to NS1 vs. HCV-1

J1 HCV-1										
	Pro Leu Val T CCG CTC GTC A --- --- ---									
J1	11	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala	Arg
		GGC	GCC	CCC	TTA	GGG	GGC	GCT	GCC	AGG
		---	---	--T	C-T	--A	---	---	---	---
J1	38	Ala	Leu	Ala	His	Gly	Val	Arg	Val	Leu
		GCC	CTG	GCA	CAT	GGT	GTC	CGG	GTT	CTG
		---	---	--G	---	--C	---	---	---	---
J1	65	Glu	Asp	Gly	Val	Asn	Tyr	Ala	Thr	Gly
		GAG	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG
		--A	---	---	---	---	---	---	---	---
J1	92	Asn	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile
		AAT	TTG	CCC	GGT	TGC	TCT	TTC	TCT	ATC
		--C	C-T	--T	---	---	---	---	---	---
J1	119	Phe	Leu	Leu	Ala	Leu	Leu	Ser	Cys	Leu
		TTC	CTC	TTG	GCT	CTG	CTG	TCC	TGT	TTG
		---	--T	C--	---	---	--C	--T	--C	---
J1	146	Thr	Ile	Pro	Ala	Ser	Ala	Tyr	Glu	Val
		ACC	ATC	CCA	GCT	TCC	GCT	TAT	GAA	GTG
		--T	G-G	--C	---	--G	--C	--C	C--	---
			Val						Gln	
J1	173	Arg	Asn	Val	Ser	Gly	Ile	Tyr	His	Val
		CGC	AAC	GTG	TCC	GGG	ATA	TAC	CAT	GTC
		---	---	TCC	A-G	---	C-T	---	--C	---
				Ser	Thr		Leu			
J1	200	Thr	Asn	Asp	Cys	Ser	Asn	Ser	Ser	Ile
		ACA	AAC	GAC	TGC	TCC	AAC	TCA	AGC	ATT
		--C	--T	--T	---	C-T	---	--G	--T	---
						Pro				

FIG. 14-1

J1	227	Val Tyr Glu Ala Ala Asp Val Ile Met GTG TAT GAG GCG GCG GAC GTG ATC ATG --- --C --- --- --C --T -CC --- C-- Ala Leu
J1	254	His Ala Pro Gly Cys Val Pro Cys Val CAT GCC CCC GGG TGC GTG CCC TGC GTT --C A-T --G --- --- --C --T --- --- Thr
J1	281	Arg Glu Asn Asn Ser Ser Arg Cys Trp CGG GAG AAC AAT TCC TCC CGT TGC TGG --T --- GG- --C G-- --G A-G --T --- Gly Ala
J1	308	Val Ala Leu Thr Pro Thr Leu Ala Ala GTA GCG CTC ACT CCC ACG CTC GCG GCC --A --- A-G --C --T --- G-G --C A-- Met Val Thr
J1	335	Arg Asn Ala Ser Val Pro Thr Thr Thr AGG AAT GCC AGC GTC CCC ACT ACG ACA --- G-- -G- -AA C-- --- G-G --- CAG Asp Gly Lys Leu Ala
Gln		
J1	362	Leu Arg Arg His Val Asp Leu Leu Val TTA CGA CGC CAC GTC GAC TTG CTC GTT C-T --- --T --- A-- --T C-- --T --C Ile
J1	389	Gly Thr Ala Ala Phe Cys Ser Ala Met GGG ACG GCT GCT TTC TGC TCC GCT ATG --- -GC --C A-C C-- --T --G --C C-C Ser Thr Leu Leu
J1	416	Tyr Val Gly Asp Leu Cys Gly Ser Val TAC GTG GGG GAT CTC TGC GGA TCT GTT --- --- --- --C --A --- --G --- --C

FIG. 14-2

J1	443	Phe Leu Ile Ser Gln Leu Phe Thr Phe
		TTC CTC ATC TCC CAG CTG TTC ACC TTC
		--T --T G-- GG- --A ---- --
		Val Gly
J1	470	Ser Pro Arg Arg His Glu Thr Val Gln
		TCG CCT CGC CGG CAT GAG ACA GTA CAG
		--T --C A-G --C --C TG- --G ACG --A
		Trp Thr
J1	497	Asp Cys Asn Cys Ser Ile Tyr Pro Gly
		GAC TGC AAC TGC TCA ATC TAT CCC GGC
		-GT ---- --T ---- --T ---- --
		Gly
J1	524	His Val Ser Gly His Arg Met Ala Trp
		CAC GTA TCA GGC CAT CGC ATG GCT TGG
		--T A-- A-G --T --C ---- --A ----
		Ile Thr
J1	551	Asp Met Met Met Asn Trp Ser Pro Thr
		GAT ATG ATG ATG AAC TGG TCG CCC ACG
		---- ---- ---- ---- ---- --C --T ----
J1	578	Ala Ala Leu Val Val Ser Gln Leu Leu
		GCA GCC TTA GTG GTG TCG CAG TTA CTC
		A-G --G --G --A A-- G-T ---- C-G ----
		Thr Met Ala
J1	605	Arg Ile Pro Gln Ala Val Met Asp Met
		CGG ATC CCA CAA GCT GTC ATG GAC ATG
		---- ---- ---- ---- --C A-- T-- ---- --
		Ile Leu
J1	632	Val Ala Gly Ala His Trp Gly Val Leu
		GTG GCG GGG GCC CAC TGG GGA GTC CTA
		A-C --T --T --T ---- ---- ---- --G
		Ile

FIG. 14-3

J1	659	Ala Gly Leu Ala Tyr Tyr Ser Met Val
		GCG GGC CTT GCC TAC TAT TCC ATG GTG
		--- --- A-A --G --T -TC --- --- ---
		Ile Phe
J1	686	Gly Asn Trp Ala Lys Val Leu Ile Val
		GGG AAC TGG GCT AAG GTT TTG ATT GTG
		--- --- --- --G --- --C C-- G-A ---
		Val
J1	713	Met Leu Leu Phe Ala Gly Val Asp Gly
		ATG CTA CTC TTT GCC GGC GTT GAC GGG
		C-- --G --A --- --- --- --C --- -C-
		Leu Ala
J1	740	His Thr Arg Val Thr Gly Gly Val Gln
		CAT ACC CGC GTG ACG GGG GGG GTG CAA
		G-A --- -A- --C --C --- --A AGT GCC
		Glu His Ser Ala
J1	767	Gly His Val Thr Ser Thr Leu Thr Ser
		GGC CAC GTC ACC TCT ACA CTC ACG TCC
		--- --- ACT GTG --- GGA T-T GTT AG-
		Thr Val Gly Phe Val
J1	794	Leu Phe Arg Pro Gly Ala Ser Gln Lys
		CTC TTT AGA CCT GGG GCG TCC CAG AAA
		--- C-C GC- --A --C --C AAG C-- --C
		Leu Ala Lys Asn
J1	821	Ile Gln Leu Val Asn Thr Asn Gly Ser
		ATT CAG CTT GTA AAC ACC AAT GGC AGT
		G-C --- --G A-C --- --- --C --- ---
		Val Ile
J1	848	Trp His Ile Asn Arg Thr Ala Leu Asn
		TGG CAT ATC AAC AGG ACT GCC CTG AAC
		--- --C C-- --T --C --G --- --- ---
		Ser

FIG. 14-4

J1	875	Cys Asn Asp Ser Leu Gln Thr Gly Phe
		TGC AAT GAC TCC CTC CAA ACT GGG TTC
		---- --T AG- --- A-C --C --C -GG
		Asn Trp
J1	902	Leu Ala Ala
		CTT GCC GCG CTG TTC TAC ACA CAC AAG
		T-G --A -G- --T --- --T CAC C-- ---
		Gly His
J1	929	TTC AAC GCG TCC GGA TGC CCG GAG CGC
		---- --T-T --A --C --T --T --- A-G
		Ser
J1	956	ATG GCC AGC TGT CGC TCC ATT GAC AAG
		C-A --- --- --C --A C-- C-- AC- G-T
		Leu Pro Leu Thr Asp
J1	983	Phe Asp Gln Gly Trp Gly Pro Ile Thr
		TTC GAC CAG GGA TGG GGT CCC ATC ACC
		--T --- --- --C --- --C --T --- -GT
		Ser
J1	1010	Tyr Ala Gln Pro Asp Asn Ser Asp Gln
		TAT GCT CAA CCT GAC AAC TCG GAC CAG
		--- --C AAC GGA AGC GG- C-C --- ---
		Asn Gly Ser Gly Pro
J1	1037	Arg Pro Tyr Cys Trp His Tyr Ala Pro
		AGG CCG TAT TGC TGG CAC TAC GCA CCT
		C-C --C --C --- --- --- --- C-C --A
		Pro
J1	1064	Arg Gln Cys Gly Ile Val Pro Ala Ser
		CGA CAG TGT GGT ATC GTA CCC GCG TCG
		AA- -CT --C --- --T --G --- --- AA-
		Lys Pro Lys

FIG. 14-5

J1	1091	Gln Val Cys Gly Pro Val Tyr Cys Phe CAG GTG TGC GGT CCA GTG TAT TGC TTC AGT --- --T --- --G --A --- --- Ser
J1	1118	Thr Pro Ser Pro Val Val Val Gly Thr ACC CCA AGC CCT GTT GTA GTG GGG ACG --T --C --- --C --G --G --- --A ---
J1	1145	Thr Asp Arg Phe Gly Ala Pro Thr Tyr ACC GAT CGT TTC GGC GCC CCT ACG TAT --- --- A-G -CG --- --G --C --C --C Ser
J1	1172	Asn Trp Gly Asp Asn Glu Thr Asp Val AAC TGG GGG GAC AAT GAG ACG GAC GTG -G- --- --T --A --- --T --- --- --C Ser Glu Asp
J1	1199	Leu Leu Leu Asn Asn Thr Arg Pro Pro CTG CTC CTA AAC AAC ACG CGG CCC CCG T-C G-- --T --- --T --C A-- --A --- Phe Val
J1	1226	His Gly Asn Trp Phe Gly Cys Thr CAC GGC AAC TGG TTC GGC TGT ACA -TG --- --T --- --- --T --- -- Leu

FIG. 14-6

J1	1	Gly Asn Trp Phe Gly Cys Thr Trp Met
HCV-1		TG GGC AAC TGG TTC GGC TGT ACA TGG ATG
		-- ---- --T ---- --T ---- --C ----
J1	30	Asn Ser Thr Gly Phe Thr Lys Thr Cys
HCV-1		AAT AGC ACT GGG TTC ACC AAG ACG TGC
		--C TCA ---- --A ---- --A GT- ---
		Val
J1	57	Gly Gly Pro Pro Cys Asn Ile Gly Gly
HCV-1		GGA GGC CCC CCG TGT AAC ATC GGG GGG
		---- -CG --T --T ---- GT- ---- --A ----
		Val
J1	84	Val Gly Asn Asn Thr Leu Thr Cys Pro
HCV-1		GTC GGC AAC AAC ACC TTG ACC TGC CCC
		-CG ---- ---- ---- ---- C-- CA- ----
		Ala His
J1	111	Thr Asp Cys Phe Arg Lys Thr Pro Thr
HCV-1		ACG GAC TGC TTC CGG AAG ACC CCG ACG
		--T --T ---- ---- --C ---- CAT --- GAC
		His Asp
J1	138	Ala Thr Tyr Thr Lys Cys Gly Ser Gly
HCV-1		GCC ACT TAC ACA AAA TGT GGT TCG GGC
		---- --A ---- T-T CGG --C --C --C --T
		Ser Arg
J1	165	Pro Trp Leu Thr Pro Arg Cys Leu Val
HCV-1		CCT TGG TTG ACA CCT AGG TGC TTG GTT
		--C ---- A-C ---- --C ---- --C --C
		Ile
J1	192	Asp Tyr Pro Tyr Arg Leu Trp His Tyr
HCV-1		GAC TAC CCA TAC AGG CTC TGG CAC TAC
		---- ---- --G --T ---- --T ---- --T --T

FIG. 15-1

J1	219	Pro	Cys	Thr	Val	Asn	Phe	Thr	Ile	Phe
HCV-1		CCC	TGC	ACT	GTC	AAC	TTT	ACC	ATC	TTC
		--T	--T	--C	A--	---	-AC	---	--A	--T
					Ile		Tyr			
J1	246	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val
HCV-1		AAG	GTT	AGG	ATG	TAT	GTG	GGG	GGC	GTG
		--A	A-C	---	---	--C	---	--A	--G	--C
			Ile							
J1	273	Glu	His							
HCV-1		GAG	CAC							
		--A	---							

FIG. 15-2

C200 region sequence vs. HCV-1

C200						3799		Asn Met Ser	
HCV-1	3781	ACA	CTG	GGC	TTT	GGT	GCT	AAT ATG TCC	
		Thr	Leu	Gly	Phe	Gly	Ala	T-C --- ---	
								Tyr	
C200	3808	Lys	Ala	His	Gly	Thr	Asp	Pro Asn Ile	
HCV-1		AAG	GCA	CAT	GGC	ACC	GAC	CCC AAC ATC	
		---	--T	---	--G	-T-	--T	--T --- ---	
						Ile			
C200	3835	Arg	Thr	Gly	Val	Arg	Thr	Ile Thr Thr	
HCV-1		AGA	ACT	GGG	GTA	AGG	ACC	ATC ACC ACA	
		--G	--C	---	--G	--A	--A	--T --- --T	
C200	3862	Gly	Ala	Pro	Ile	Thr	Tyr	Ser Thr Tyr	
HCV-1		GGT	GCC	CCC	ATT	ACG	TAC	TCC ACC TAT	
		--C	AG-	---	--C	---	---	---	--C
			Ser						
C200	3889	Arg	Lys	Phe	Leu	Ala	Asp	Gly Gly Cys	
HCV-1		CGC	AAG	TTC	CTT	GCC	GAC	GGT GGT TGC	
		G--	---	---	---	---	---	--C --G ---	
		Gly							
C200	3916	Ser	Gly	Gly	Ala	Tyr	Asp	Ile Ile	
HCV-1		TCC	GGG	GGC	GCC	TAT	GAC	ATC ATA A	
		--G	---	---	--T	---	---	--A --- -TT	
								Ile	
HCV-1	3943	TGT	GAC	GAG	TGC	CAC	TCC	ACG GAT GCC	
		Cys	Asp	Glu	Cys	His	Ser	Thr Asp Ala	
HCV-1	3970	ACA	TCC	ATC	TTG	GGC	ATC	GGC ACT GTC	
		Thr	Ser	Ile	Leu	Gly	Ile	Gly Thr Val	

FIG. 16-1

HCV-1	3997	CTT GAC CAA GCA GAG ACT GCG GGG GCG	Leu Asp Gln Ala Glu Thr Ala Gly Ala
HCV-1	4024	AGA CTG GTT GTG CTC GCC ACC GCC ACC	Arg Leu Val Val Leu Ala Thr Ala Thr
HCV-1	4051	CCT CCG GGC TCC GTC ACT GTG CCC CAT	Pro Pro Gly Ser Val Thr Val Pro His
HCV-1	4078	CCC AAC ATC GAG GAG GTT GCT CTG TCC	Pro Asn Ile Glu Glu Val Ala Leu Ser
HCV-1	4105	ACC ACC GGA GAG ATC CCT TTT TAC GGC	Thr Thr Gly Glu Ile Pro Phe Tyr Gly
C200	4132	Ser Ile Pro Ile Glu Ala Ile Lys	
HCV-1		A AGC ATC CCC ATC GAG GCC ATC AAG	
		AAG GCT --- --- C-- --A -TA --- ---	
		Lys Ala	Val
C200	4159	Gly Gly Arg His Leu Ile Phe Cys His	
HCV-1		GGG GGA AGG CAT CTC ATC TTC TGC CAT	
		--- --G --A --- --- --- --- --T ---	
C200	4186	Ser Lys Lys Lys Cys Asp Glu Leu Ala	
HCV-1		TCC AAG AAG AAG TGT GAC GAG CTC GCC	
		--A --- --- --- --C --- --A --- ---	
C200	4213	Ala Lys Leu Ser Ala Leu Gly Leu Asn	
HCV-1		GCA AAG CTG TCA GCC CTC GGA CTC AAT	
		--- --- --- GTC --A T-G --C A-- ---	
			Val Ile
C200	4240	Ala Val Ala Tyr Tyr Arg Gly Leu Asp	
HCV-1		GCC GTG GCG TAT TAC CGC GGT CTT GAT	
		--- --- --C --C --- --- --- --- --C	

FIG. 16-2

C200	4267	Val	Ser	Val	Ile	Pro	Thr	Ser	Gly	Asp
HCV-1		GTG	TCC	GTC	ATA	CCA	ACT	AGC	GGA	GAC
		---	---	---	--C	--G	--C	---	--C	--T
C200	4294	Val	Val	Val	Val	Ala	Thr	Asp		
HCV-1		GTC	GTT	GTC	GTG	GCA	ACA	GAC	GC	4316
		--T	--C	---	---	---	--C	--T	--C	CTC
HCV-1	4321	ATG	ACC	GGC	TAT	ACC	GGC	GAC	TTC	GAC
		Met	Thr	Gly	Tyr	Thr	Gly	Asp	Phe	Asp
HCV-1	4348	TCG	GTG	ATA	GAC	TGC	AAT	ACG	TGT	GTC
		Ser	Val	Ile	Asp	Cys	Asn	Thr	Cys	Val
HCV-1	4375	ACC	CAG	ACA	GTC	GAT	TTC	AGC	CTT	GAC
		Thr	Gln	Thr	Val	Asp	Phe	Ser	Leu	Asp
HCV-1	4402	CCT	ACC	TTC	ACC	ATT	GAG	ACA	ATC	ACG
		Pro	Thr	Phe	Thr	Ile	Glu	Thr	Ile	Thr
HCV-1	4429	CTC	CCC	CAG	GAT	GCT	GTC	TCC	CGC	ACT
		Leu	Pro	Gln	Asp	Ala	Val	Ser	Arg	Thr
HCV-1	4456	CAA	CGT	CGG	GGC	AGG	ACT	GGC	AGG	GGG
		Gln	Arg	Arg	Gly	Arg	Thr	Gly	Arg	Gly
HCV-1	4483	AAG	CCA	GGC	ATC	TAC	AGA	TTT	GTG	GCA
		Lys	Pro	Gly	Ile	Tyr	Arg	Phe	Val	Ala
HCV-1	4510	CCG	GGG	GAG	CGC	CCC	TCC	GGC	ATG	TTC
		Pro	Gly	Glu	Arg	Pro	Ser	Gly	Met	Phe
HCV-1	4537	GAC	TCG	TCC	GTC	CTC	TGT	GAG	TGC	TAT
		Asp	Ser	Ser	Val	Leu	Cys	Glu	Cys	Tyr

FIG. 16-3

HCV-1 4564 GAC GCA GGC TGT GCT TGG TAT GAG CTC
 Asp Ala Gly Cys Ala Trp Tyr Glu Leu

HCV-1 4591 ACG CCC GCC GAG ACT ACA GTT AGG CTA
 Thr Pro Ala Glu Thr Thr Val Arg Leu

HCV-1 4618 CGA GCG TAC ATG AAC ACC CCG GGG CTT
 Arg Ala Tyr Met Asn Thr Pro Gly Leu

HCV-1 4645 CCC GTG TGC CAG GAC CAT CTT GAA TTT
 Pro Val Cys Gln Asp His Leu Glu Phe

HCV-1 4672 TGG GAG GGC GTC TTT ACA GGC CTC ACT
 Trp Glu Gly Val Phe Thr Gly Leu Thr

HCV-1 4699 CAT ATA GAT GCC CAC TTT CTA TCC CAG
 His Ile Asp Ala His Phe Leu Ser Gln

HCV-1 4726 ACA AAG CAG AGT GGG GAG AAC CTT CCT
 Thr Lys Gln Ser Gly Glu Asn Leu Pro

HCV-1 4753 TAC CTG GTA GCG TAC CAA GCC ACC GTG
 Tyr Leu Val Ala Tyr Gln Ala Thr Val

HCV-1 4780 TGC GCT AGG GCT CAA GCC CCT CCC CCA
 Cys Ala Arg Ala Gln Ala Pro Pro Pro

HCV-1 4807 TCG TGG GAC CAG ATG TGG AAG TGT TTG
 Ser Trp Asp Gln Met Trp Lys Cys Leu

HCV-1 4834 ATT CGC CTC AAG CCC ACC CTC CAT GGG
 Ile Arg Leu Lys Pro Thr Leu His Gly

HCV-1 4861 CCA ACA CCC CTG CTA TAC AGA CTG GGC
 Pro Thr Pro Leu Leu Tyr Arg Leu Gly

FIG. 16-4

HCV-1	4888	GCT	GTT	CAG	AAT	GAA	ATC	ACC	CTG	ACG	
		Ala	Val	Gln	Asn	Glu	Ile	Thr	Leu	Thr	
HCV-1	4915	CAC	CCA	GTC	ACC	AAA	TAC	ATC	ATG	ACA	
		His	Pro	Val	Thr	Lys	Tyr	Ile	Met	Thr	
HCV-1	4942	TGC	ATG	TCG	GCC	GAC	CTG	GAG	GTC	GTC	
		Cys	Met	Ser	Ala	Asp	Leu	Glu	Val	Val	
HCV-1	4969	ACG	AGC	ACC	TGG	GTG	CTC	GTT	GGC	GGC	
		Thr	Ser	Thr	Trp	Val	Leu	Val	Gly	Gly	
HCV-1	4996	GTC	CTG	GCT	GCT	TTG	GCC	GCG	TAT	TGC	
		Val	Leu	Ala	Ala	Leu	Ala	Ala	Tyr	Cys	
HCV-1	5023	CTG	TCA	ACA	GGC	TGC	GTG	GTC	ATA	GTG	
		Leu	Ser	Thr	Gly	Cys	Val	Val	Ile	Val	
HCV-1	5050	GGC	AGG	GTC	GTC	TTG	TCC	GGG	AAG	CCG	
		Gly	Arg	Val	Val	Leu	Ser	Gly	Lys	Pro	
C200									Glu	Val	Leu
									GAA	GTC	CTC
HCV-1	5077	GCA	ATC	ATA	CCT	GAC	AGG	---	---	---	
		Ala	Ile	Ile	Pro	Asp	Arg				
		Tyr	Arg	Glu	Phe	Asp	Glu	Met	Glu	Glu	
C200	5104	TAC	CGA	GAG	TTC	GAT	GAG	ATG	GAA	GAG	
HCV-1		---	---	---	---	---	---	---	---	---	
		Cys	Ala	Ser	His	Leu	Pro	Tyr	Ile	Glu	
C200	5131	TGC	GCC	TCA	CAC	CTC	CCC	TAC	ATC	GAA	
HCV-1		---	T-T	CAG	---	T-A	--G	---	---	--G	
			Ser	Gln							

FIG. 16-5

C200	5158	Gln Gly Met Gln Leu Ala Glu Gln Phe
HCV-1		CAG GGA ATG CAG CTC GCC GAG CAA TTC
		--A --G --- AT- --- --- --G ---
		Met
C200	5185	Lys Gln Lys Ala Leu Gly Leu Leu Gln
HCV-1		AAG CAG AAG GCG CTC GGG TTG CTG CAA
		--- --- --- --C --- --C C-C --- --G
C200	5212	Thr Ala Thr Lys Gln Ala Glu Ala Ala
HCV-1		ACA GCC ACC AAG CAA GCG GAG GCT GCT
		--C --G T-- CGT --G --A --- -T- ATC
		Ser Arg Val Ile
C200	5239	Ala Pro Cys Glu Ser Met His Ala Ser
HCV-1		GCT CCG TGT GAG TCA ATG CAC GCC TCG
		--C --T GC- -TC CAG -CC A-- TGG CAA
		Ala Val Gln Thr Asn Trp Gln
C200	5266	A
HCV-1		-AA CTC GAG ACC TTC TGG GCG AAG CAT
		Lys Leu Glu Thr Phe Trp Ala Lys His
HCV-1	5293	ATG TGG AAC TTC ATC AGT GGG ATA CAA TA
		Met Trp Asn Phe Ile Ser Gly Ile Gln

FIG. 16-6

NS1 Sequence vs. HCV-1

J1 HCV-1	1	Leu Gly Asn Trp Phe Gly Cys Thr Trp G TTG GGC AAT TGG TTC GGT TGC ACC TGG - C-- --- --- --- --- --- ---T --- ---
J1 HCV-1	29	Met Asn Ser Ser Gly Phe Thr Lys Val ATG AAC TCA TCT GGA TTT ACC AAA GTG --- --- --- A-- --- --C --- --- --- Thr
J1 HCV-1 Ala	56	Cys Gly Ala Pro Pro Cys Val Ile Gly TGC GGA GCG CCT CCT TGT GTC ATC GGA --- --- --- --- --- --- --- --- ---
J1 HCV-1	83	Gly Val Gly Asn Asn Thr Leu Gln Cys GGG GTG GGC AAC AAC ACC TTG CAA TGC --- -C- --- --- --- C-- --C --- Ala His
J1 HCV-1	110	Pro Thr Asp Cys Phe Arg Lys His Pro CCC ACT GAC TGT TTC CGC AAG CAT CCG --- --- --T --- --- --- --- --- ---
J1 HCV-1	137	Asp Ala Thr Tyr Ser Arg Cys Gly Ser GAC GCC ACA TAC TCT CGG TGC GGT TCC --- --- --- --- --- --- --- --C ---
J1 HCV-1	164	Gly Pro Trp Ile Thr Pro Arg Cys Leu GGT CCC TGG ATT ACG CCC AGG TGC CTG --- --- --- --C --A --- --- --- ---
J1 HCV-1	191	Val His Tyr Pro Tyr Arg Leu Trp His GTC CAC TAC CCT TAT AGG CTT TGG CAT --- G-- --- --G --- --- --- --- --- Asp
J1 HCV-1	218	Tyr Pro Cys Thr Val Asn Tyr Thr Leu TAT CCC TGT ACT GTC AAC TAC ACC TTG --- --T --- --C A-- --- --- --- A-A Ile Ile

FIG. 17-1

J1	245	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly
HCV-1		TTC	AAA	GTC	AGG	ATG	TAC	GTG	GGA	GGG
		--T	---	A--	----	----	----	----	----	----
				Ile						
J1	272	Val	Glu	His	Arg	Leu	Glu	Val	Ala	Cys
HCV-1		GTC	GAG	CAC	AGG	CTG	GAA	GTT	GCT	TGC
		----	--A	----	----	----	----	-C-	--C	----
								Ala		
J1	299	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	Asp
HCV-1		AAC	TGG	ACG	CGG	GGC	GAG	CGT	TGT	GAT
		----	----	----	----	----	--A	----	--C	----
J1	326	Leu	Asp	Asp	Arg	Asp				
HCV-1		CTG	GAC	GAC	AGG	GAC	A			
		----	--A	----	----	----	--			
			Glu							

FIG. 17-2

Core Sequence vs. HCV-1

J1 HCV-1	1	GCGTCTAGCCATGGCGTTAGTATGAGTGTC -----
J1 HCV-1	31	GTGCAGCCTCCAGGACCCCCCTCCCGGGAGAGCC -----
J1 HCV-1	66	ATAGTGGTCTGCGGAACCGGTGAGTACACCGGAAT -----
J1 HCV-1	101	TGCCAGGACGACCGGGTCCTTTCTTGGATCAACCC -----
J1 HCV-1	136	GCTCAATGCCTGGAGATTTGGGCGTGCCCCCGCGA -----A-
J1 HCV-1	171	GACTGCTAGCCGAGTAGTGTGGGTGCGGAAAGGC -----
J1 HCV-1	206	CTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGC -----
J1 HCV-1	241	CCCGGGAGGTCTCGTAGACCGTGTCATCATG AGC -----C-----Met Ser
J1 HCV-1	274	Thr Asn Pro Lys Pro Gln Arg Lys Thr ACA AAT CCT AAA CCT CAA AGA AAA ACC --G --- --- --- --- -A- --- -A- Lys Asn
J1 HCV-1	301	Lys Arg Asn Thr Asn Arg Arg Pro Gln AAA CGT AAC ACC AAC CGC CGC CCA CAG --- --- --- --- --- -T --- ---
J1 HCV-1	328	Asp Val Lys Phe Pro Gly Gly Gly Gln GAC GTC AAG TTC CCG GGC GGT GGT CAG --- --- --- --- --- -T -C ---
J1 HCV-1	355	Ile Val Gly Gly Val Tyr Leu Leu Pro ATC GTT GGT GGA GTT TAC CTG TTG CCG --- --- --- --- --- T-- ---
J1 HCV-1	382	Arg Arg Gly Pro Arg Leu Gly Val Arg CGC AGG GGC CCC AGG TTG GGT GTG CGC --- --- --- -T -A --- --- ---

FIG. 18-1

J1 HCV-1	409	Ala Thr Arg Lys Thr Ser Glu Arg Ser GCG ACT AGG AAG ACT TCC GAG CGG TCG --- --G --A --- --- --- --- --- ---
J1 HCV-1	436	Gln Pro Arg Gly Arg Arg Gln Pro Ile CAA CCT CGT GGA AGG CGA CAA CCT ATC --- --- --A --T --A --T --G --- ---
J1 HCV-1	463	Pro Lys Ala Arg Gln Pro Glu Gly Arg CCC AAG GCT CGC CAG CCC GAG GGC AGG --- --- --- --T -G- --- --- --- --- Arg
J1 HCV-1	490	Ala Trp Ala Gln Pro Gly Tyr Pro Trp GCC TGG GCT CAG CCC GGG TAC CCT TGG A-- --- --- --- --- --- --- --- --- Thr
J1 HCV-1	517	Pro Leu Tyr Gly Asn Glu Gly Met Gly CCC CTC TAT GGC AAC GAG GGC ATG GGG --- --- --- --- --T --- --- TGC --- Cys
J1 HCV-1	544	Trp Ala Gly Trp Leu TGG GCA GGA TGG CTC CT --- --G --- --- --- --

FIG. 18-2